AAMRL-TR-88-024

THE ROLE OF SHORT-TERM MEMORY IN OPERATOR WORKLOAD (U)

AD-A200 252

DTIC
ELECTE
OCT 1 2 1988
S D
D

WILLIAM F. REINHART
CARITA D. GLYNN, M.S.
CRAIG DYE
M. TAKAHAMA, M.S.
HARRY L. SNYDER, Ph.D.

VIRGINIA POLYTECHNIC INSTITUTE AND STATE UNIVERSITY

APRIL 1988

FINAL REPORT FOR JANUARY 1987 - APRIL 1988

Approved for public release; distribution is unlimited.

ARMSTRONG AEROSPACE MEDICAL RESEARCH LABORATORY
HUMAN SYSTEMS DIVISION
AIR FORCE SYSTEMS COMMAND
WRIGHT-PATTERSON AIR FORCE BASE, OHIO 45433-6573

88 1011 277

NOTICES

TECHNICAL REVIEW AND APPROVAL

AAMRL-TR-88-024

This report has been reviewed by the Office of Public Affairs (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS, it will be available to the general public, including foreign nations.

This tecnical report has been reviewed and is approved for publication.

FOR THE COMMANDER

CHARLES BATES, JR.
Director, Human Engineering Division
Armstrong Aerospace Medical Research Laboratory

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704-0188

1a. REPORT SECURITY CLASSIFICATION
UNCLASSIFIED

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION / DOWNGRADING SCHEDULE

3. DISTRIBUTION / AVAILABILITY OF REPORT
Approved for public release; distribution is unlimited.

4. PERFORMING ORGANIZATION REPORT NUMBER(S)

5. MONITORING ORGANIZATION REPORT NUMBER(S)
AAMRL-TR-88-024

6a. NAME OF PERFORMING ORGANIZATION
Virginia Polytechnic Institute and State University*

6b. OFFICE SYMBOL *(If applicable)*

7a. NAME OF MONITORING ORGANIZATION
AAMRL/HED

6c. ADDRESS *(City, State, and ZIP Code)*
Displays and Controls Laboratory
302 Whittemore Hall
Blacksburg VA 24061

7b. ADDRESS *(City, State, and ZIP Code)*
Wright-Patterson AFB OH 45433-6573

8a. NAME OF FUNDING / SPONSORING ORGANIZATION

8b. OFFICE SYMBOL *(If applicable)*

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
F33615-85-D-0514 (0016)

8c. ADDRESS *(City, State, and ZIP Code)*

10. SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT ACCESSION NO.
62202F	7184	10	3C

11. TITLE *(Include Security Classification)*
The Role of Short-Term Memory in Operator Workload (U)

12. PERSONAL AUTHOR(S)
Reinhart, William F., Glynn, Carita D., Dye, Craig, Takahama, M. and Snyder, Harry L.

13a. TYPE OF REPORT	13b. TIME COVERED	14. DATE OF REPORT (Year, Month, Day)	15. PAGE COUNT
Final	FROM Jan 87 TO Apr 88	1988 April	177

16. SUPPLEMENTARY NOTATION
*Contract awarded through Southeastern Center for Electrical Engineering Education, Central Florida Facility, 1101 Massachusetts Ave., St. Cloud, Florida 32769

17. COSATI CODES

FIELD	GROUP	SUB-GROUP
05	08	
05	09	

18. SUBJECT TERMS *(Continue on reverse if necessary and identify by block number)*
Advanced Conceptual Bomber, Crew Station Design, Memory, Mental Workload, Modified Cooper-Harper Scale, Short-Term Memory, Sternberg Scanning Paradigm, Subjective Workload Assessment Technique, Workload

19. ABSTRACT *(Continue on reverse if necessary and identify by block number)*
The role of the flight crew member in the advanced conceptual bomber is likely to call for rapid management of complex information. The elevated workload associated with these new informational tasks may be alleviated in part by the reduction of short-term memory demands. Literature reviews were conducted to described the concepts of short-term memory and workload. An experiment was conducted to describe the contribution of short-term memory demands to operator workload. The Sternberg scanning model was used in a secondary task paradigm with various performance and subjective measures recorded. Sternberg choice reaction time and subjective ratings of workload increased as a function of visual memory set size. It was concluded that short-term memory may significantly contribute to operator workload. Therefore strategies for the reduction of short-term memory demands in crew station information display design should be investigated.

20. DISTRIBUTION / AVAILABILITY OF ABSTRACT
☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS

21. ABSTRACT SECURITY CLASSIFICATION
UNCLASSIFIED

22a. NAME OF RESPONSIBLE INDIVIDUAL
Mr. Gilbert G. Kuperman

22b. TELEPHONE *(Include Area Code)*
(513) 255-3727

22c. OFFICE SYMBOL
AAMRL/HEA

DD Form 1473, JUN 86 *Previous editions are obsolete.*

SECURITY CLASSIFICATION OF THIS PAGE
UNCLASSIFIED

PREFACE

This research was performed under contract No. F33615-85-D-0514 for the Harry G. Armstrong Aerospace Medical Research Laboratory (AAMRL). The principal aim of this project is to provide experimental groundwork and validity for the use of the short-term memory concept in the design of workload-reducing bomber information management technologies. Sections 1 and 2 contain brief literature reviews on the subjects of short-term memory and mental workload. Section 3 describes preliminary research performed at this laboratory. Section 4 contains a summary of the initial experiment and recommendations for future research.

The authors thank Mr. Gilbert Kuperman and Ms. Denise Wilson for their guidance in the preparation of this report. They also thank Mr. Gary Reid for providing the conjoint analysis software which was used in the SWAT analysis.

TABLE OF CONTENTS

TABLE OF CONTENTS (continued)

TABLE OF CONTENTS (continued)

LIST OF TABLES

LIST OF ILLUSTRATIONS

I. SHORT-TERM MEMORY LITERATURE REVIEW

INTRODUCTION

The so called "biological barrier" refers to system performance limitations which stem from sensory and behavioral characteristics of the human operator in the system loop. It is a commonly discussed problem in the venue of aviation research and development, where rapidly evolving flight system and aerial combat technologies have pushed aircraft performance envelopes to and beyond the capacity of the human aircrewman. While much attention has been paid to this problem in the context of the one or two-man fighter aircraft, relatively little discussion has been centered on the advanced manned bomber.

Increased complexity of threat system technology, and consequently increased manned bomber mission complexity, point to a likely increase in mental workload demands for bomber aircrews (Kuperman and Wilson, 1986). Because weapon system flexibility, effectiveness, and survivability may be jeopardized by such elevated workload, enhancement of the crewmember-machine interface (CMI) has gained added importance for the design of the advanced manned bomber crew station.

Thompson (1981) predicted that the role of a flight crew member will evolve from one of a "flyer" to a "flight information manager." Likewise, Gopher (1982) has pointed to the new role of pilots as "monitors and supervisors of numerous, rapidly changing flight systems" (p. 173). If these predictions are true, then advanced information management strategies and task structuring must be considered for implementation in the advanced bomber environment with a careful weighting of those elements which significantly contribute to aircrew mental workload. One such important element may be short-term memory.

There is ample evidence from both the short-term memory and workload literature to suggest that these two concepts are vitally linked and may be fruitfully employed together. The area of fault detection in automated systems serves as a good example. Curry (1981) stated one of the fundamental assumptions in the implementation of automated systems: "There is a [workload] cost to monitoring which can be alleviated by use of [machine] monitoring systems" (p. 175). Having stated this assumption, Curry then said, "The amount of information available to the [human monitor] will depend on [in part] the short-term memory capacity" (p. 176). Wickens and Kessel (1981) also concluded that human operator characteristics (e.g., short-term memory) are vital to the assessment of workload in failure detection tasks.

Another example may be found in aviation psychology. Aviation psychologists have found that some short-term memory tasks closely mirror naturally occurring information processing tasks in typical flight profiles (Loftus, Dark, and Williams, 1979). Experimental manipulations of information processing rates and retention intervals have pointed to a link between short-term memory failure and pilot communications errors. Furthermore, discussion may be found in the literature for implementation of short-term memory tasks as indices of pilot mental workload (e.g., Wickens, Hyman, Dellinger, Taylor, and Meador, 1986).

This literature review was undertaken to determine the state of contemporary research on the subject of human short-term memory. To date, no one model of short-term memory has provided a definitive description of the complex concept of human memory. However, certain trends in the development of both methodologies and models afford some answers and clues as to the functional nature of short-term memory and its role in information processing.

This section, along with the review of the mental workload literature (Section II), serves as a foundation in the design and implementation of an experimental battery for the quantification of the short-term memory/workload relationship.

Contents

The remainder of this review is divided into five parts. In the first part, traditionally used research methodologies in the study of human short-term memory are illustrated. In the second part, a number of task and stimulus variables which have been shown to influence measures of short-term memory are presented. A description of the development and status of various models of memory may be found in the third part. Part 4 contains a discussion of possible strategies to be used in the reduction of short-term memory demands or the extention of short-term memory processing capabilities. In the fifth part, the role of short-term memory in mental workload and various applied environments is outlined.

RESEARCH METHODOLOGY

Introduction

It has long been recognized that experimental memory data are to a large degree a product of the paradigm used to generate them (Craik and Lockhart, 1972) (see Testing Paradigms). It is equally true that paradigmatical details must be tailored to specific research questions if the resulting memory data are to have any meaning. While certain memory paradigms provide better conceptual fits than others, there is no such thing as a general purpose short-term memory paradigm. Thus, the following research strategies represent only frameworks and

examples from which to draw a more detailed and precise strategy for hypothesis testing. All have received extensive use, in a variety of forms, so that familiarity with these examples in the literature will provide further explication of the possible paradigmatical variations, anomalies, and families of curves.

Memory research is traditionally differentiated between recall and recognition paradigms. Recall paradigms require subjects to reproduce some stimulus or group of stimuli from memory upon the experimenter's cue. Recall data are typically reported in terms of error percentage or span of correct recall (the number of correctly recalled stimuli). On the other hand, recognition paradigms simply require the acknowledgement by the subject that a stimulus or group of stimuli has been seen before and encoded in memory. Error percentage or choice reaction time are typical dependent variables in the recognition memory paradigm.

Recall paradigms

Free Recall. When subjects are asked to recall items from a stimulus group without the restrictions of sequence or position found in other recall paradigms, they are performing free recall tasks. In a free recall task, subjects are simply asked to recall as many items from the original stimulus set as possible. Stimuli may be presented in series or simultaneously, usually allowing no more than a few seconds per item (Underwood, 1983). Pacing of recall may also be used, as in the use of the metronome by Peterson and Peterson (1959). The number of correct responses, rather than the error rate, usually serves as the dependent variable.

The free recall paradigm has been popular largely due to its utility in studying the recency effect (Tulving, 1968). The

recency effect is the empirical observation that items last presented in a serial list show a higher proportion of recall than those items serially previous. Free recall tasks were predominant in the literature during the structuralist movement between the 1950s and 1970s (Wingfield and Byrnes, 1981), with the recency effect having been interpreted by many as evidence for the existence of a separate, short-term memory store.

Serial Recall. Crannell and Parrish (1957) used a serial recall paradigm to investigate differences in short-term memory span for digits, letters, and words. Serial recall involves the recall of stimuli in the serial order of original presentation. It is also known as ordered or immediate recall (Puff, 1982). Crannell and Parrish found digits to yield the highest percent correct responses for all lists used, while word stimuli elicited the poorest performance.

As in the free recall task, the number of correct responses is usually taken as the dependent variable in the serial recall paradigm. In cases where single trial stimulus presentation is used before recall (as by Crannell and Parrish, 1957) performance is frequently expressed in terms of memory span.

Whole versus Partial Report. Recall tasks are typically whole report tasks. That is, subjects are asked to recall as many items as possible from the original stimulus set. In 1960, Sperling noted that the memory span reported in whole report recall tasks may be confounded with a time constraint on subjects' ability to complete their reports (due to decay of the memory trace). That is, in the time required to complete a verbal report, significant memory decay may occur. Consequently, recall scores for long stimulus sets may not accurately reflect the instantaneous short-term memory capacity of a subject at the time he is asked to begin his report. This effect has been interpreted as evidence for a channel capacity

limitation in the recognition memory buffer, which has properties of "fast read-in and slow read-out" (Bundesen, Pedersen, and Larsen, 1984, p. 329).

In response to this whole report confound, Sperling invented the partial report task. In the partial report task, subjects are asked to report only part of the information presented to them. Sperling argued that by selectively sampling from positions in the original stimulus set (partial report), experimenters should be able to predict the difference between short-term memory span and that amount of memory which has been lost during the interval of report. Figure 1 shows obtained memory spans from Sperling's 1960 experiment. Estimates from the partial report technique place the immediate memory span for 12 stimulus letters at almost two times the number found in the whole report condition, suggesting a large time-to-report constraint in the whole report paradigm.

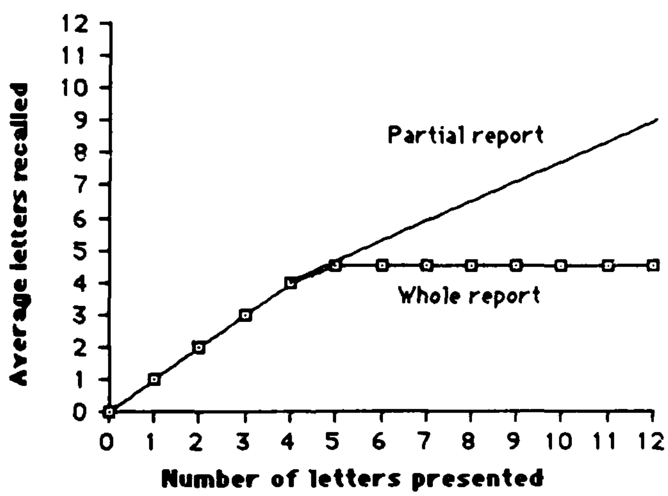


Figure 1. Whole Report Span Versus Partial Report Estimate, After Sperling (1960).

Distractor Tasks. Distractor tasks are commonly known as Peterson-Brown tasks, after their founders. Peterson and Peterson (1959) and Brown (1958) introduced the use of the distractor task to prevent subject rehearsal in recall tasks. In the research of Peterson and Peterson (1959), the distractor task involved counting backwards by threes or fours at a constant rate (a metronome was used) for varying lengths of time. At the presentation of a signal, subjects were to recall (in serial order) the three consonants and the three digit number which was presented to them previous to the distractor task. The results (Figure 2) show a steady decline in performance as a function of time spent performing the distractor task.

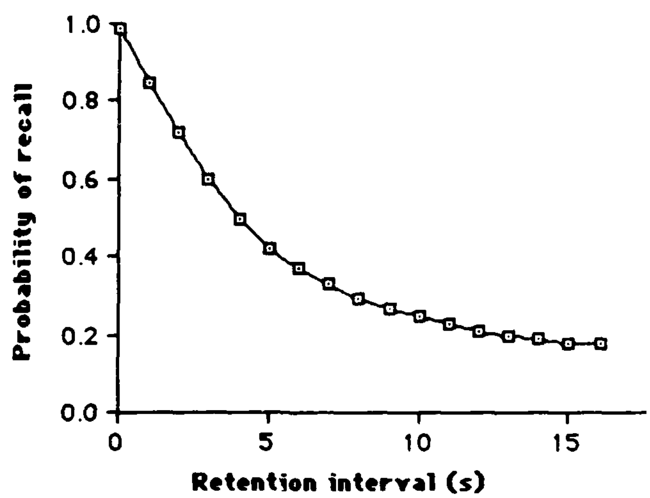


Figure 2. Recall Performance in a Distractor Paradigm, After Peterson and Peterson (1959).

Subsequently Posner and Rossman (1965) demonstrated that when the interval of delay was held constant, distractor task difficulty maintained a systematic reduction of recall performance (Puff, 1982). These data seem to rule out a strict time-decay interpretation and suggest a possible cognitive resource competition interpretation.

Probe Recall. Unlike serial recall tasks which require subjects to recall information in the order in which it was presented, the probe recall paradigm samples particular elements from the stimulus set. Probe recall tasks may be either sequential, position, or paired-associates probes (Puff, 1982). Waugh and Norman (1965) provided an example of the sequential probe task, in which subjects were presented a probe digit (one of 16 previously presented digits) and were asked to recall the digit which had immediately followed it.

The position probe task requires recall of a single stimulus element from a particular order or position in a list or spatial location. For example, a subject might be asked to recall the fourth letter from a previously presented list of ten letters. Atkinson and Shiffrin (1968) used the position probe in testing recall of the color of cards positioned in a spatial sequence.

In the paired-associates probe task, a number of pair-associates are presented. For example, pairs of color names such as the words "green" and "blue" might be presented. Next, one of the two elements from a pair (e.g., "green") is presented and subjects are asked to report the identity of the missing element from that pair (i.e., "blue"). In the unidirectional mode, response B is performed to probe A. In a bidirectional paradigm, the subject's response is required to either element of the pair (Underwood, 1983). The research of Murdock (1963) serves as an example. Murdock presented

subjects with paired English words. After six trials, one of the words was presented again in solicitation of its pair word. As is common in recall tasks, presentation rate was varied, in this case among one, two, and three seconds per pair. Multiple probes can also be used as a variation of this paradigm.

Release from Proactive Inhibition. Proactive inhibition is the tendency for previously learned information to interfere with the learning of new, but similar information. Proactive inhibition may accumulate over a series of distractor trials (Keppel and Underwood, 1962). Since switching to trials of a different stimulus class alleviates the proactive inhibition effect on recall performance, Wickens, Born, and Allen (1963) reasoned that this "release from proactive inhibition" technique could yield an index of the differences among various stimulus classes by demonstrating differential release effects. The release from proactive inhibition effect is well documented and can be achieved by changing a wide range of stimulus dimensions (Baddeley, 1982). For example, Allen (1984) produced release from proactive inhibition when his subjects switched from learning color names to visual colors.

Recognition Paradigms

Differential Probe. As in the probe recall paradigm, after the study period in which the stimulus set is introduced, a probe is presented. Subjects are asked to judge whether the probe has membership in the original stimulus set. Half the probe items are usually new items (not of the original stimulus set). Typically, the percentage of correct responses is recorded, although subjective confidence ratings and latency of response have also been used (e.g., Sternberg, 1966). Shulman (1970) used a differential probe task to study semantic coding in short-term memory. His subjects were required to recognize whether the probe was identical, homonomous, or synonomous with

one of 10 list words.

Sternberg Scanning. Latency of response is the primary measure of subject performance in the Sternberg Scanning Paradigm (Sternberg, 1966, 1969 a). This paradigm is really a subclass of the differential probe task. Subjects are presented with small stimulus sets (usually one to six items) and then given a recognition probe test. Subjects must decide if the probe is a member of the original stimulus set (MSET). Subjects respond "yes" or "no" as quickly as possible and choice reaction time is recorded. Trials may be either fixed, using the same MSET for numerous probes, or varied, with probes presented only once for each MSET.

Choice reaction time is important because Sternberg's model assumes a serial and exhaustive search of short-term memory (Puff, 1982), with an increase in search time as the length of MSET increases. The function of reaction time plotted against MSET size (Figure 3) yields a slope which is inversely related to capacity of working memory (Cavanaugh, 1972) and efficiency of memory search (Wickens et al., 1986). That is, since in a function with a low slope each additional MSET item adds relatively little to total response time, the capacity and efficiency of short-term memory in this case is interpreted as being high.

By varying the size of MSET, the class of stimulus material, and the probability of response direction (positive or negative), "the total processing time may be broken down into a time to encode the stimulus, a time to scan the memory, ...and a time to select and execute the response" (Wickens et al., 1983, p. 1372). Figure 3 shows mean response latencies from Sternberg's 1966 experiment.

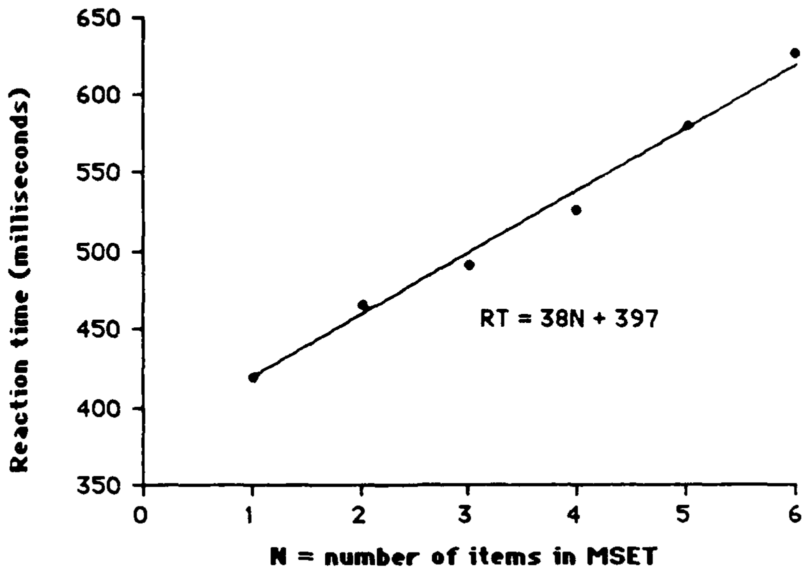


Figure 3. Mean Response Latencies for Eight Subjects and Six Values of MSET Size, After Sternberg (1966).

This function can be expressed as:

$$RT = CN + (e + d), \tag{1}$$

where e is the time in ms required to read and encode the probe digit, C is the time in ms required to scan one item, N is the number of items in MSET, and d is the time in ms necessary to arrive at a decision and execute a response (Loftus and Loftus, 1976). Expressed as in formula (1), RT is a linear function of N with slope C and an intercept of (e + d). More simply, the slope reflects the efficiency of scanning or working memory, while input/output delays are reflected in the intercept of the function.

The second important measure used with the Sternberg task is percentage error. This must be held to a relatively low rate since the interpretability of the RT data depends upon a successful memory scan (Wickens et al., 1986). Toward this end, MSET sizes used should be subspan, that is, below what the expected maximum short-term memory span would be (approximately 7 +/- 2 items).

In Sternberg's data, the slope (the scan time per digit) was calculated to be 38 ms per item. This time has been shown to vary with other stimuli such as words (Chase and Calfee, 1969) and random forms (Sternberg, 1969 b). Also, negative responses are consistently slower than positive responses, although this difference remains constant across various MSET sizes (Wickens et al., 1986).

Cavanaugh (1972) has shown a reciprocal relationship to exist between processing rate and memory span. Figure 4 shows processing rate in ms/item on the ordinate. These values are obtained as slopes in linear functions such as that shown in Figure 3. High processing rates indicate inefficient or slow memory scans. The abscissa in Figure 4 is the reciprocal of memory span. A large reciprocal value indicates a low memory span. It can be seen in Figure 4 that memory span and processing rate vary together as a function of stimulus class.

Signal Detection. Murdock (in Puff, 1982) described the use of a signal detection analysis in the study of recognition memory. The advantage of this application includes the use of one summary statement of recognition accuracy (d') instead of two (performance on old and new items) as well as the means of separating decision criteria from recognition performance.

A signal detection treatment of recognition memory assumes two

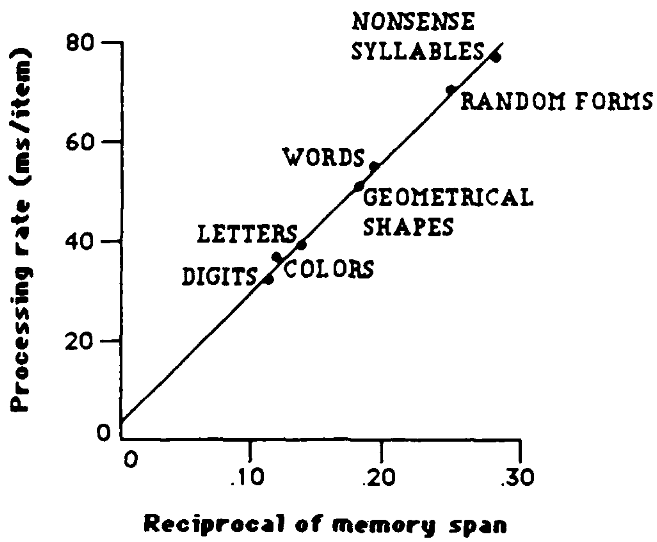


Figure 4. Short-Term Memory Processing Rate and the Reciprocal of Memory Span, After Cavanagh (1972).

---

overlapping distributions on the memory trace strength continuum (Figure 5). One distribution represents the variable strength of the trace for old items (those from the original stimulus list) and one distribution represents the trace strength for new items, introduced during the probe phase. These distributions are usually assumed to be normal and of equal variance.

A criterion point may be located on the memory trace strength continuum corresponding to the point at which a recognition probe will be called by the subject "new" or "old." Because memory strength is variable, and because the two distributions usually overlap, a certain proportion of the responses to the right of the criterion will be false alarms. Likewise, a

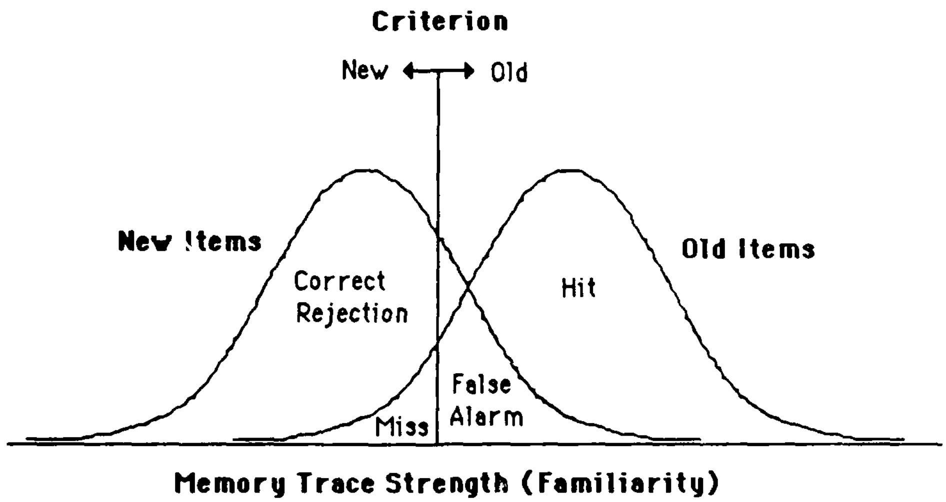


Figure 5. A Hypothetical Distribution of New and Old Memory Items.

---

proportion of the "new" responses to the left of the criterion will be misses. The balance of the responses will be either hits or correct rejections. A measure of subject response sensitivity, d', may be obtained as the difference between the two standardized means.

Table 1 shows hypothetical confidence judgements to new and old probe items for a probe recognition task. The use of such confidence measures allows the collection of several data points in one session, since each level of confidence is interpreted to represent a separate level of response criteria. These frequencies can be converted to hit and false alarm rates by computing the cumulative probability of hits and false alarms at each criterion level minus one (Table 2). Hit and false alarm rates may then be used to construct a memory operating characteristic (MOC) curve. The MOC curve is analogous to the ROC curve in other signal detection

TABLE 1. Frequencies of Six Confidence Judgements to New and Old Memory Items, from Murdock, in Puff (1982).

	Confidence Judgements					
	---	--	-	+	++	+++
Old items	25	35	40	40	28	32
New items	90	50	28	18	10	4

TABLE 2. Hit and False Alarm Rates from Data in Table 1, from Murdock, in Puff (1982).

Criterion	Hits	False Alarms
+++	.16	.02
++	.30	.07
+	.50	.16
-	.70	.30
--	.88	.55

procedures.

Figure 6 shows the MOC curve for the data in Table 2. The MOC curve reflects the effect of a change in response criterion on the probability of hits and false alarms, as a function of subject sensitivity (d'). A higher d' is shown by a more pronounced bow to the curve.

Variations

Set Span. Recall and recognition paradigms which are used to investigate memory spans may use either subspan or supraspan stimulus sets. A subspan set is one with fewer elements than would normally define the maximum span of memory in a given paradigm. A classic example of the use of the subspan list is found in Sternberg's scanning paradigm. In order to meet the assumption of the scanning model, it is necessary to obtain

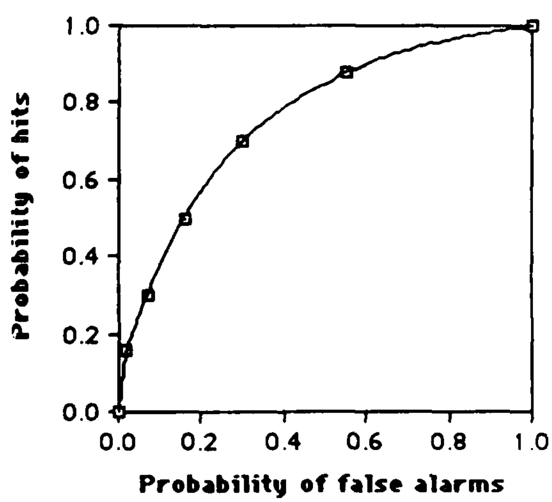


Figure 6. The MOC Curve for the Data in Table 2, after Murdock, in Puff (1982).

choice RT scores that are near to error-free. For this reason, subspan lists are used which are just below the expected maximum span of short-term memory. Supraspan sets have a greater number of elements than would normally define the maximum span of memory. Supraspan lists have traditionally been used to study accuracy and to quantify the maximum retention span of material. For example, Crannell and Parrish (1957) used supraspan lists in their comparisons of digit, letter, and word spans.

Loading. Another variant in the presentation of stimuli occurs in the use of memory loads or distractor tasks. The Peterson-Brown distractor task has already been discussed and involves the imposition of an irrelevant task during the retention interval. In contrast, preloads and concurrent memory loads occur before and during stimulus presentation.

Baddeley and Hitch (1974) provided an example of preloading. Baddeley and Hitch's subjects were presented with a digit list prior to presentation of a word list. Recall was later required for both lists. Preloading had a deleterious effect on the primary, but not on the recent portion of the free recall curves for the word lists.

Concurrent loading was reported in an experiment by Baddeley, Grant, Wight, and Thomson (1975). A pursuit tracking task was performed by subjects during presentation of paired-associate word lists. Concurrent loading significantly impaired recall of paired-associate words.

CRITICAL VARIABLES IN SHORT-TERM MEMORY

Introduction

There are no doubt many operator and extra-task variables which may affect the function of human short-term memory. Such factors as fatigue, motivation, and environmental conditions may modify memory performance within specific contexts, but will not be considered here.

Factors to be considered in this section are largely task and stimulus specific. For example, the role of rehearsal is one that is integral to most of the models of memory considered in this report. The importance of rehearsal can also be seen in the extensive use of distractor tasks to disrupt rehearsal and so arrive at a better understanding of short-term processes. In addition to other task properties, stimulus variables such as modality, semantic meaningfulness, and novelty will be considered here as important factors affecting the function of short-term memory.

Rehearsal

Rehearsal interference (e.g., the Peterson-Brown distractor task) has been widely used to demonstrate the role of rehearsal in the coding of information from short-term storage to long-term storage. Rehearsal has also been modeled to have a role in the maintenance of information in short-term memory. It has thus been assumed by proponents of a duplex model of memory that the duration of retention in short-term memory is a function of the amount of rehearsal time available (Atkinson and Shiffrin, 1968; Waugh and Norman, 1965). However, Craik and Watkins (1973) measured short-term storage times and found under some conditions no reliable prediction of either long-term recall or recognition as a function of time in

storage. Craik and Watkins suggested two separate rehearsal roles may be involved: a maintenance rehearsal system which holds information in short-term store, and an elaborative rehearsal system which facilitates long-term encoding.

The bulk of the literature in this area reports use of verbal material and the role of articulatory rehearsal (e.g., Baddeley, Thomson, and Buchanan, 1975). Recent literature, however, has addressed the role of visual rehearsal as well. These studies indicate that rehearsal of visuo-spatial materials plays a role in short-term retention and long-term encoding similar to that of the articulatory loop (e.g., Baddeley and Lieberman, 1980). Differential effects of modalities will be discussed more fully in the following passages.

Stimulus Modality

Evidence to suggest the division of the short-term store into modality specific mechanisms or channels is abundant. Posner and Keele's (1967) report supported the existence of a visuo-spatial information store. In their study, "same" judgements were made faster for identical stimulus pairs (eg., AA) than for visually different pairs (eg., Aa).

Baddeley and Lieberman (1980) discriminated between visual and spatial memory. They supported their model by citing selective interference in the performance of a spatial memory task by a secondary spatial tracking task. A secondary visual tracking task produced no such interference.

Baddeley et al. (1975) showed that visual tracking performance was impaired by requiring subjects to process a visual memory image. However, impairment was not evident when the processing task was a verbal one.

A further complication in the delineation of visuo-spatial memory is that serial position curves for recognition of complex pictures usually show no recency effect (Schaffer and Shiffrin, 1972). This finding suggests a limitation in short-term storage of complex visual stimuli (Hitch, 1983).

Shapiro and Erdelyi (1974) and Erdelyi and Becker (1974) reported experiments in which hypermnesia (incrementally improved recall) was demonstrated for pictures but not for words. Unfortunately, because instructions were inserted between stimulus presentation and recall, their results cannot necessarily be generalized to short-term memory.

When verbal stimuli are presented auditorily rather that visually, greater retention and recall accuracy usually result. Wickens, Sandry, and Vidulich (1983) referred to this phenomenon as the auditory memory effect. This effect is well documented (Baddeley, 1982; Nilsson, Ohlsson, and Ronnberg, 1977). In addition, short-term serial recall performance is disrupted by phonemic similarity among list items (Baddeley, 1966; 1984).

In a replication by Allen (1984) of a visual color and color name recall task, subjects showed release from proactive inhibition when the stimulus class was shifted from color names to visual colors. However, no release was found when the shift was made from colors to color names. This unidirectional pattern of release from proactive inhibition has been obtained both with subject vocalization (Allen, 1983) and without (Allen, 1984).

In a study of short-term memory for the duration of movements, Elliot and Jones (1984) suggested that visual input interferes with the mental rehearsal of spatial information, perhaps due

to a high attentional bias toward visual information.

Semantic Meaningfulness

Crannell and Parrish (1957) presented subjects with auditory lists of digits, letters, and words. Letter and word lists were either limited to 9 possible elements (equal to the digit list pool) or were unlimited (26 possible elements). Percent correct recall was highest for digits and lowest for words. The effect of limiting letter and word pool size was not statistically significant. The authors suggested that these differences may be due in part to the relative frequency of experience with which subjects have had practice in grouping these classes of stimuli (see Grouping).

Lavach (1973) tested the effect of high and low arousal producing words in paired-associate recall. Words such as "kiss", "exam", and "love" produced high GSR arousal levels and subsequently low recall scores for short-term retention. Low arousal producing words elicited low GSR arousal levels and high short-term memory recall scores. These results suggest that low arousal conditions during stimulus acquisition foster superior short-term recall.

Finally, Baddeley (1966) used a serial recall procedure to test the effect of semantic similarity on recall of adjectives (e.g., high, tall, wide, broad). Although recall interference for semantically similar words was not as great as for acoustically similar words, there was a significant impairment of recall (6.3 % below control).

Testing Paradigms

It has long been accepted within the framework of a duplex model of memory that the recency effect and memory span in free

recall reflect the same limited capacity store (i.e., a unitary short-term memory store (Hitch, 1985)). However, a number of differential task-dependent effects have led some to challenge this interpretation.

For example, performance in immediate serial recall (from which memory span measures are typically obtained) is disrupted by factors such as phonemic similarity of stimuli (Baddeley, 1966; 1984), simultaneous digit processing (Klapp and Philipoff, 1983), simultaneous, irrelevant articulation (articulatory suppression) (Fitzgerald and Broadbent, 1985; Hitch, 1985). At the same time, none of these factors alter the recency effect in the free recall paradigm.

Finally, as outlined in the previous discussion of research methodology, partial report tasks may reflect immediate visual memory more accurately than whole report tasks (Sperling, 1960). Bundesen, Pedersen, and Larsen (1984) demonstrated superior partial report recall for selection by brightness, alphanumeric characters, and color. In addition, partial report superiority increased as the ratio of distractor items to targets increased, and decreased with a decreased ratio.

Rate of Stimulus Presentation and Processing

A number of studies have been reported which support the position that differences in memory span vary as a function of not only total storage space, but rather as the operational efficiency with which information is processed (Daneman and Carpenter, 1980; Dempster, 1981).

Case, Kurland, and Goldberg (1982) reported four studies in which storage space was defined through free recall tasks (i.e., memory span) and operational efficiency was measured as the total processing speed in separate, reaction time recall

paradigms. Case et al. demonstrated that when operational efficiency was held constant, memory span differences become insignificant. These results support the implication that differences in memory span are attributable to differences in operational efficiency. They further suggested that rate of stimulus presentation may be as important as, if not more important than, stimulus set length in determining recall performance.

In a related study, Ellis and Hennelly (1980) demonstrated that differences in the amount of time necessary to articulate Welsh and English digits accounted for differences in digit span scores for Welsh and English children.

Baddeley et al. (1975) were able to predict immediate memory span by the number of words read in two seconds (i.e., subject reading speed). Two seconds is also the time at which inverted response (reponse in the opposite order of stimulus input) purportedly yields the greatest improvement in immediate auditory recall (Posner, 1964).

Finally, McKendry and Hurst (1971) demonstrated the effects of exceeding subject channel capacity for rate of visual information input. They concluded that such speed stress can be adapted to through practice. As evidence, they cited faster response times and lower error rates following practice exposures to speed stress. Not surprisingly, speed stress thresholds were lower for more complex stimuli.

MODELS OF HUMAN MEMORY

Introduction

The popular understanding of human memory is probably best represented in a model proposed by Atkinson and Shiffrin (1971) (Figure 7). This model earned the title of "modal model" (Baddeley, 1984) precisely because of the widespread inroads it made in both popular and scientific circles. This model, an elaboration of the simple duplex model of human memory proposed by Waugh and Norman (1965), still serves as groundwork for today's models. Although it continues to dominate the popular understanding of memory, this simple model no longer enjoys complete support in the literature today. In fact, the duplex concept of memory (the idea that short-term and long-term memory are two distinct intraorganismic stores) has never been without its critics (see Melton, 1963). Since many of the models to be discussed here assume a duplex foundation, it is important to first consider some of the evidence for and against the duplex position.

As discussed in RESEARCH METHODOLOGIES, some of the most compelling evidence to suggest two functionally distinct memory stores has been generated in free recall experimentation. In the free recall paradigm, stimuli are presented to the subject, after which he must recall as many items as possible from the stimulus set. Although stimuli may be presented simultaneously, when they are presented sequentially it is possible to generate a serial position curve. The serial position curve reflects the proportion of correct item recall relative to the position of the item in the original stimulus sequence. A hypothetical serial position curve is shown in Figure 8.

In a typical free recall experiment, the probability of recall

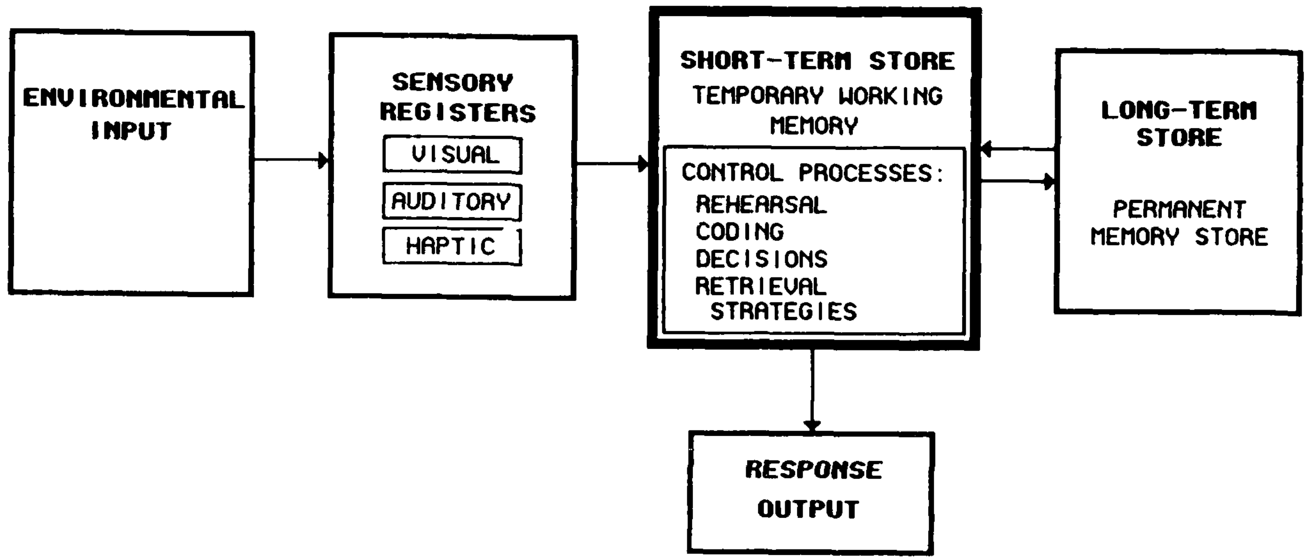


Figure 7. Atkinson and Shiffrin's "Modal Model," after Atkinson and Shiffrin (1971).

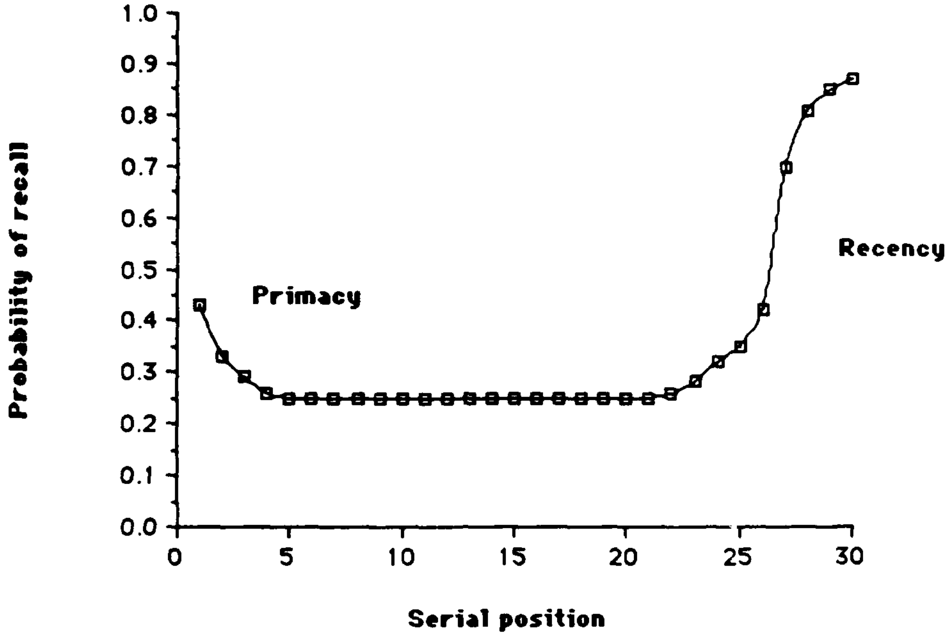


Figure 8. A hypothetical Serial Position Curve.

---

tends to be the highest for the most recently presented stimuli (the recency effect), the next highest for the earliest trial (the primacy effect), and fairly monotonic in between (Craik, 1970). The primacy effect is interpreted as evidence for a long-term storage facility. Presumably, these items have had a longer time interval in which to become encoded into a long-term store through the process of rehearsal. Early list items, then, should have a higher probability of recall than intermediate and later items. This long-term memory store fails, however, to by itself account for the recency effect.

The popular interpretation of the recency effect leads to the proposition of a short-term memory store. That is, since recent items have the least amount of time in which to be

rehearsed, and yet show the highest probability of recall, another store is suggested which serves to hold the memory trace prior to the transmittal to long-term storage.

Further evidence for a short-term memory store comes from the introduction of the distractor task paradigm. By preventing rehearsal, experimenters have been able to document a progressive recall decrement as a function of time. A typical retention curve of this nature (Figure 8) shows a dramatic decrease in probability of recall in as little as five seconds. This momentary preservation of the short-term component, without later evidence of long-term retrieval, is interpreted as further evidence for the duplex model of memory and the role of rehearsal in long-term encoding.

Finally, cases of selective loss of short-term memory, inability to form new long-term memory (anterograde amnesia), and sometimes both (Baddeley, 1982; Cermak, 1982) are reported in a body of clinical data from amnesiac individuals. Vallar and Baddeley (1984) presented a clinical case as evidence for the existence of an articulatory rehearsal loop, one component of the Baddeley and Hitch (1974) working memory hypothesis. Clinical drug studies (e.g., Mewaldt, Hinrichs, and Ghoneim, 1983) have added further support by showing the selective disruption of specific elements of working memory.

Distinctions between the two memory stores have been made on the basis of these data and more. The commonly discussed dimensions of functional difference are summarized in Table 3. These differences include capacity, duration or persistence, and instrumentation of information loss (forgetting).

Van der Heijden (1981) distinguished between two classes of information processing models: precategorical and postcategorical selection models. Precategorical selection models assume a limited information processing or

FEATURE	CONCEPTS	CHARACTERISTICS
Memory processes	Short-term memory (STM)	Information passed to STM, where it is held for up to 30 seconds if not rehearsed.
	Long-term memory (LTM)	Information may be stored in LTM on a more permanent basis.
Distinguishing short and long-term memory	Temporary versus relative permanence	STM is temporary, LTM is more permanent.
	Capacity	STM includes 7+/- 2 pieces of information; LTM is immense.
	Primacy/recency effects	Primacy reflects LTM; Recency reflects STM.
	Forgetting	Displacement is prominent in STM; Interference is prominent in LTM.
Processes in short-term memory	Coding	Auditory coding is primary, but imagery and semantic coding are also important.
	Retrieval	Search can occur very rapidly, and we may search each item.

TABLE 3. Differential characteristics of memory processes and capacities, after Santrock (1986).

categorization capacity due to the limitation of some selection mechanism. Precategorical selection models presented in this section include Sperling's (1963) linear information processing model and Crowder and Morton's (1969) PAS model, although the model on which the PAS is built (Morton's (1969) logogen model) could be properly considered postcategorical.

Postcategorical models are given a larger representation in this review. Postcategorical selection models emphasize the organism's limited capacity for response or memory storage. Work by Waugh and Norman (1965), Atkinson and Shiffrin (1971), Baddeley and Hitch (1974), Craik and Lockhart (1972), and Gilmartin, Newell, and Simon (1976) all are included in this section as examples of postcategorical selction models.

Precategorical Selection Models

Sperling's Model. Unlike Craik and Lockhart's levels-of-processing model, Sperling's model is primarily concerned with the passage of information from sensory stores (iconic and echoic) to behavioral reports. Sperling (1963, 1967) introduced a scanner and recognition buffer (Figure 9) between the sensory stores and short-term stores. Sperling was puzzled by the ability of subjects to report as many as five letters from a brief visual display, despite previous data which suggested that iconic traces were useful for only up to 500 ms (Gregg, 1986). Sperling reasoned that since subject report itself took longer than the duration of the trace, some mechanism must be holding the information long enough for the subject to complete the report. Sperling proposed the scanner as that mechanism.

Sperling's scanner rapidly extracts sensory information from the iconic or echoic store, encodes the information, then passes it along to the first of seven short-term storage *slots*. The short-term store also contains a rehearsal mechanism

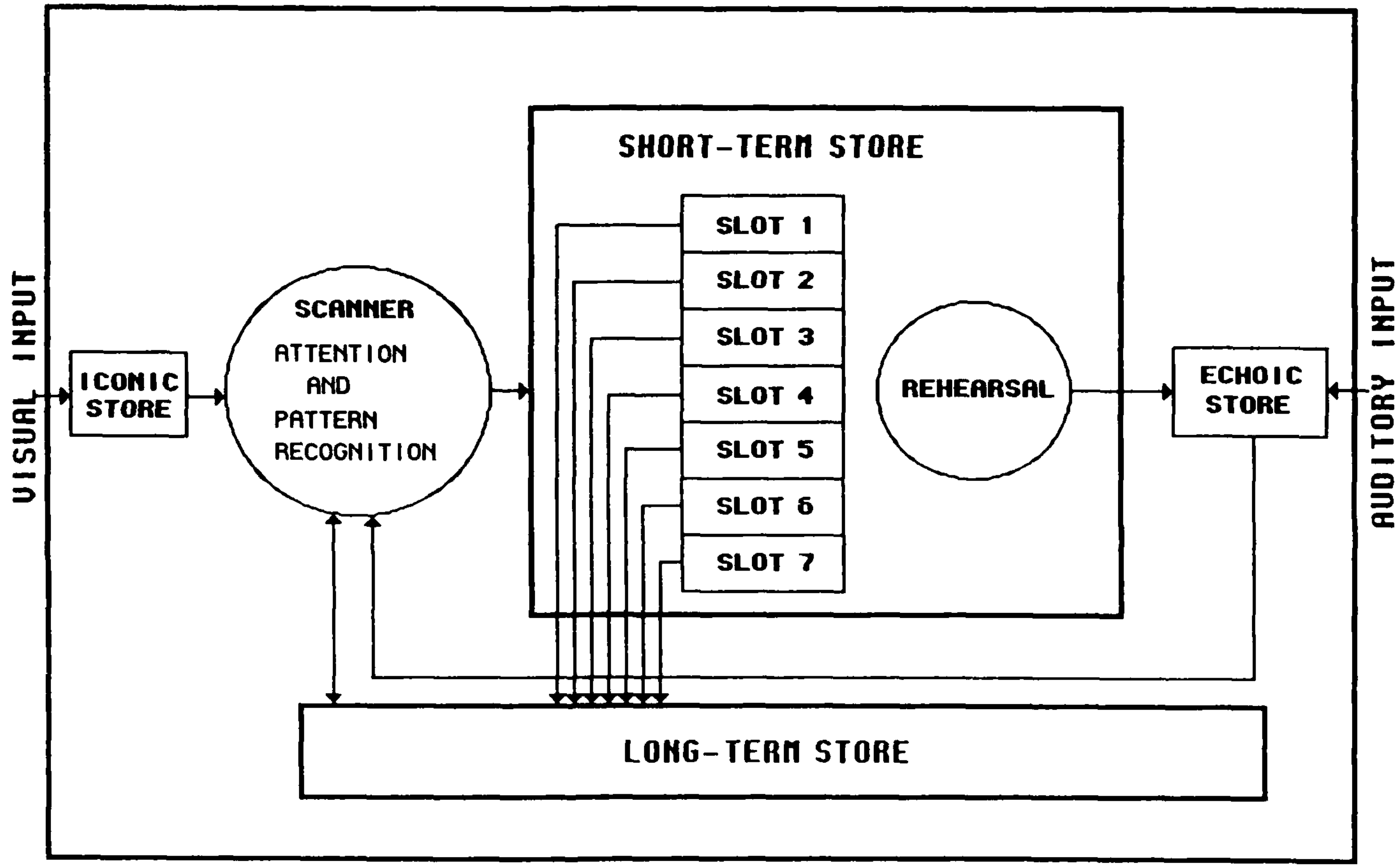


Figure 9. Sperling's (1963) information processing model, from Loftus and Loftus (1976), adapted from Sperling (1960).

outputing to the echoic store, and consequently back to the scanner, where it may once again be introduced into the short-term store. This articulatory loop is a common theme in models of memory and has received an in-depth treatment in Baddeley and Hitch's working memory model.

Crowder and Morton's PAS. The auditory suffix effect refers to the large performance decrement in auditory serial recall for the last few of eight or nine items when a redundant, not to be recalled, digit is included at the end of the digit list (Gregg, 1986). Crowder and Morton (1969) proposed a precategorical acoustic storage (PAS) unit to explain this effect.

Built on Morton's (1969) logogens model (Figure 10), the PAS is seen as a primary encoder of auditory stimuli. An analogous visual analyzer (the ICON) exists in parallel to the PAS. PAS accounts for the auditory suffix effect by suggesting that the redundant auditory digit displaces the last relevant serial digit from the limited storage facility in PAS, thus eliminating the otherwise beneficial acoustic trace present in PAS at the time of recall.

In Figure 10, the logogens is seen as the categorical buffer, where stimuli first receive categorization as verbal units (hence, "precategorical acoustic storage"). Any sensory stimulation in ICON or PAS is assumed to be retained in a more primitive code. As in most contemporary models, an articulatory rehearsal mechanism is included. This particular rehearsal loop provides explicitly for both silent and vocal rehearsal. The cognitive system is assumed to hold, among other things, the long-term storage function.

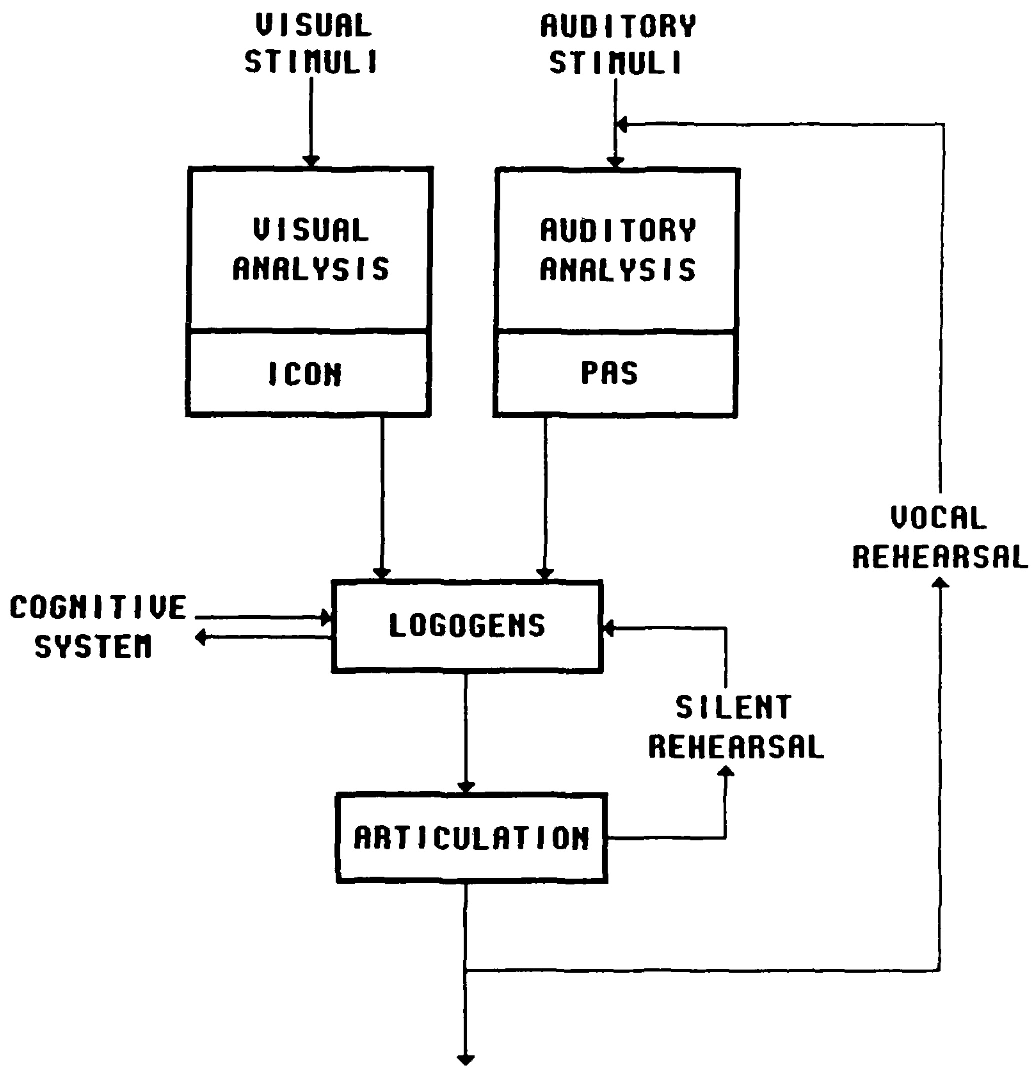


Figure 10. An early PAS model. Later models contain separate logogen units for ICON and PAS inputs. Adapted from Gregg (1986), after Crowder and Morton (1969).

Post Categorical Selection Models

Waugh and Norman's Duplex Model. Waugh and Norman (1965) borrowed the terms "primary memory" and "secondary memory" from William James (1890, cited by Gregg, 1986) for use in their model of short-term verbal retention. In the Waugh and Norman model (Figure 11), stimulus information enters primary memory where it may either be maintained and passed on to secondary memory through rehearsal, or forgotten.

While secondary memory is assumed to have unlimited storage capacity, primary memory is limited to about three words, regardless of syllable length (Craik, 1968). Primary memory and secondary memory are thus analogous to short-term and long-term stores in terms of function, capacity, and sequence of information processing.

Atkinson and Shiffrin's Model. The Atkinson and Shiffrin (1968, 1971) model also suggests separate short-term and long-term stores. However, it has several important features

---

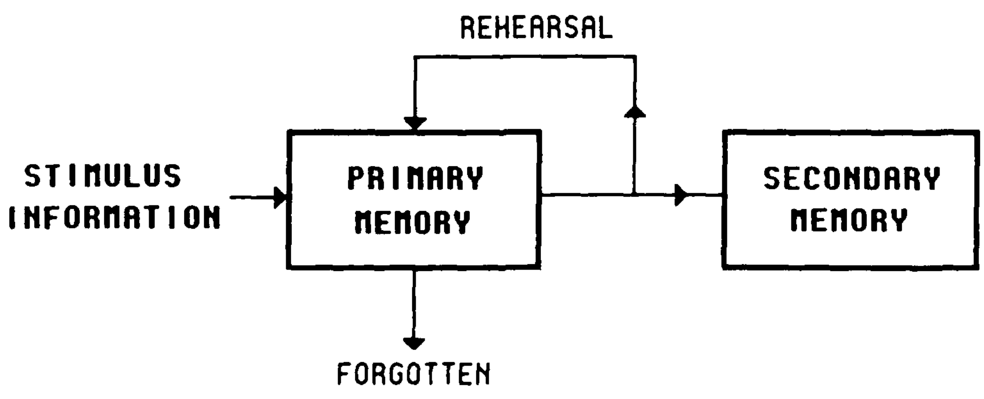


Figure 11. The Simple Duplex Model, after Waugh and Norman (1965).

which distinguish it from the Waugh and Norman model. Firstly, Atkinson and Shiffrin added sensory registers (visual, auditory, and haptic) as an intermediate process between environmental input and short-term storage (Figure 7). More importantly, however, short-term memory is seen as including not only a passive memory area, but also active resident control processes as a means of processing the contents of short-term memory. Rehearsal, coding, decision, and retrieval strategies carry out the organization, interaction with long-term storage, and response output. Again, since long-term storage is not the focus of this model, it suffers the generalized assignment of essentially unlimited and permanent storage capacity.

Working Memory. The term "working memory" is a functional description which Baddeley and Hitch (1974) have used for the role of short-term memory in information processing. Although their model assumes the duplex distinction of memory, short-term memory assumes a complexity and flexibility here which is not present in previous models such as Waugh and Norman's (1965).

Short-term memory is described as a system of secondary slave systems serving a type of central processing unit, the "central executive". Figure 12 illustrates how the central executive might be involved in the solution of an arithmetic problem (Hitch, 1978). These short-term memory subsystems were suggested by the differential effects of different types of memory loading apparent in the empirical database. For example, the "articulatory loop" is seen as a speech-based mechanism which drives subvocal rehearsal.

In a more recent development, Salame and Baddeley (1982) subdivided the articulatory loop to account for differential short-term memory disruption effects of irrelevant speech. Since articulatory suppression in auditorily presented material

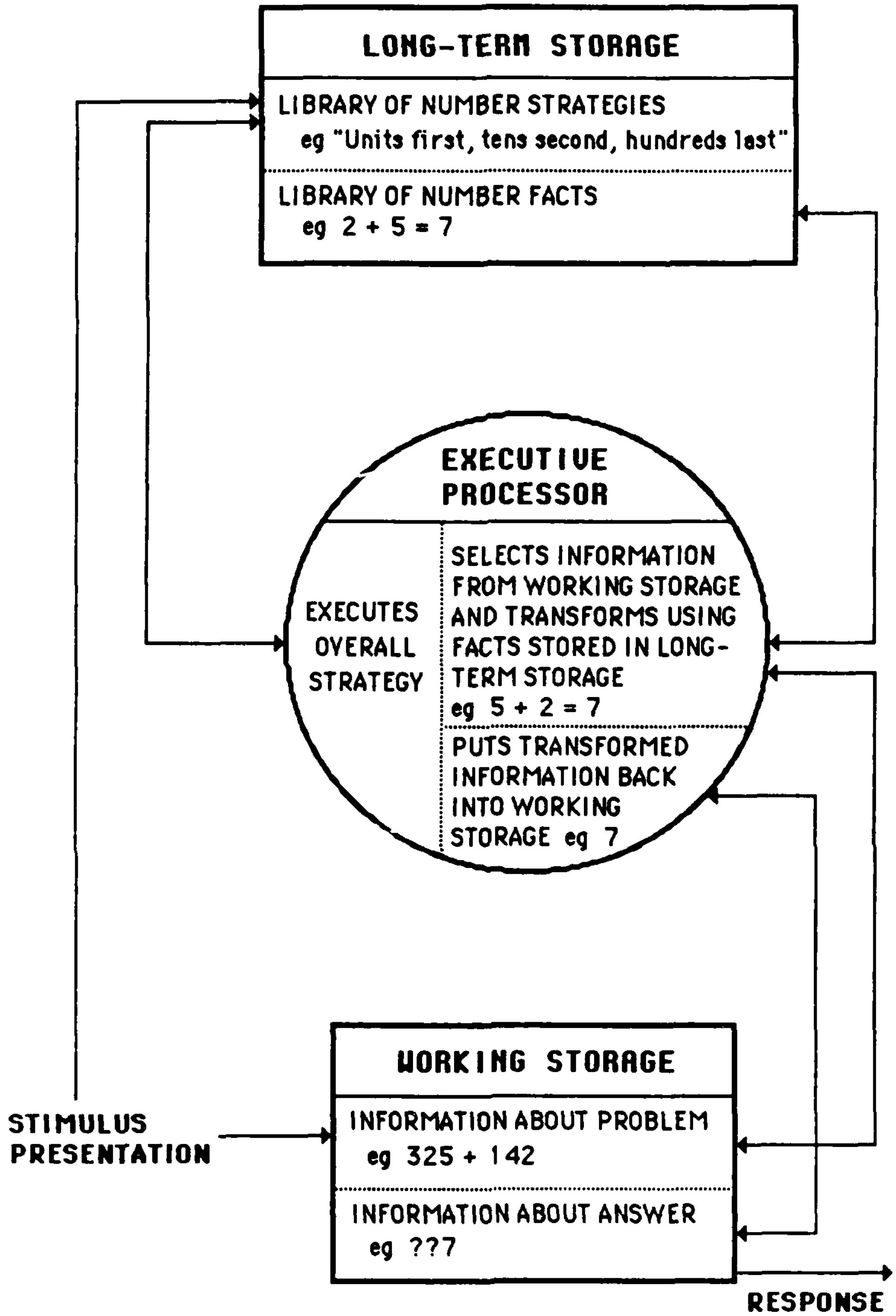


Figure 12. A structural interpretation of the role of the executive processor in an arithmetic problem, adapted from Hitch (1978).

has no effect on recall for phonetically similar words but alleviates the effect of word length in general (Baddeley, Lewis, and Vallor, cited by Hitch, 1984), Salame and Baddeley (1982) suggested the existence of a passive phonological store.

A second subsystem serving the central executive is the visuo-spatial scratch-pad (Baddeley and Leiberman, 1982). The scratch-pad is used to construct mental images and remember spatial arrangements. The visuo-spatial scratch-pad is thought to use a covert rehearsal mechanism, possibly analogous to eye movements (Hitch, 1984). According to the model, because the visuo-spatial scratch-pad timeshares the processing capacity of the central executive, a concurrent task such as mental arithmetic reduces the available processing capacity for the scratch-pad and thus interferes with the full functioning of the visual imagery system (Baddeley, 1982).

Baddeley (1982) has acknowledged that various components of the working memory model are only in the infancy of their development. The central executive in particular has received very little empirical attention relative to the articulatory loop. For example, Baddeley (1982) suggested that the central executive may itself be further subdivided to contain a primary memory unit and a mechanism for the direction of conscious attention.

Levels-of-Processing. Craik and Lockhart's levels-of-processing model was born out of a research question summarized by Craik and Watkins (1973). Their research addressed the mechanisms of transferrence of information from short-term memory into long-term memory (e.g., rehearsal). Craik and Watkins found that words held in short-term memory for long periods of time were not necessarily more likely to be recalled than those held in short-term memory for short periods of time. Craik and Lockhart proposed that short-term memory (or primary memory) must be part of a continuum in a system

capable of processing and coding information at a variety of levels. This hypothesis was further supported by the observation that some stimulus codes seem more likely than others to be encoded in long-term memory (e.g., semantically meaningful words). The levels-of-processing model holds that the depth of processing (and degree of stimulus elaboration), rather than primarily the amount of rehearsal, is the main influence in memory trace persistence.

Craik and Lockhart have advocated Moray's (1967) concept of a central information processor which directs information to various levels of analysis (Craik and Lockhart, 1972). The directed depth of analysis determines the strength of the resultant memory trace and thus the likelihood that the information will be transferred to long-term storage. Although the incorporation of a central attentional mechanism is used in a number of short-term memory models, including the Baddeley and Hitch working memory model, Craik and Lockhart's usage is considerably different. Craik and Lockhart's central executive services a hierarchy of processing levels, whereas the peripheral processes in the Baddeley and Hitch model do not necessarily imply an order of processing depth.

SHORT. Gilmartin, Newell, and Simon (1976) created a SNOBOL computer program, SHORT, which serves as an information processing model of short-term memory (Figure 13). Information enters SHORT from either the visual or auditory environment. Stimuli entering the sensory stores are held for 250 ms in the visual store and 3 s in the auditory store, in accordance with the classical data (Gilmartin et al., 1976). Perception occurs when SHORT accesses an imagery store and makes a match with previously stored patterns in long-term memory.

The short-term element of SHORT receives the product of such a perceptual match. A first-in-first-out stack array of eight cells is used. Items can be retained in short-term storage

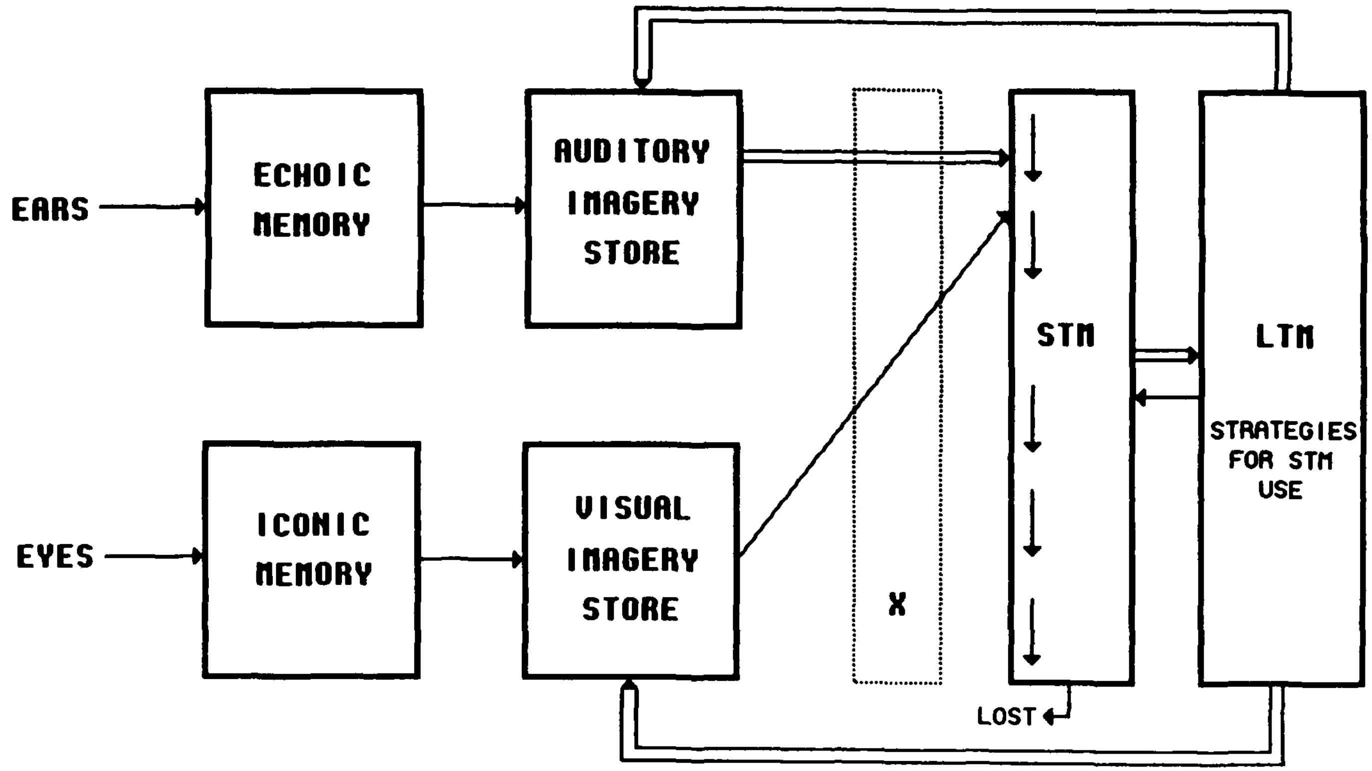


Figure 13. A representation of SHORT. Double arrows show information flow during auditory rehearsal. The "x" represents perception of encoded stimuli. After Gilmartin, Newell, and Simon (1983).

through rehearsal (imaging an item, reperceiving it, and thus moving it back to the top of the stack). Items can also be lost (pushed out the bottom of the stack by incoming items or the passage of time), or they may be transferred to long-term storage. Unlike the Atkinson and Shiffrin model (1971), SHORT places its short-term memory strategies in the long-term storage mechanism.

STRATEGIES FOR REDUCING SHORT-TERM MEMORY DEMANDS

Introduction

The literature on mnemonics (techniques or devices for improving memory) contains predominately strategies intended to aid long-term storage and retrieval. In addition, much of this literature concerns reduction coding of verbal material (e.g., Baddeley, 1976). For example, the acronym ROY.G.BIV has been used to remember the colors of the visible spectrum of electromagnetic energy (i.e., Red Orange Yellow Green Blue Indigo Violet). Visual mnemonics have been suggested as well, but again these are directed toward improvement of the long-term element. For example, the method of loci (Santrock, 1986) involves imagining a physical location for each item to be remembered.

The use of such mnemonic devices is limited in the application to long-term memory, largely due to their complicated and time consuming nature. The time required to construct an acronym for a novel set of stimuli far exceeds the likely immediate duration of that memory trace. Any virtue such a task is likely to have in relation to short-term memory is in facilitating the maintenance function of rehearsal.

Seen in this light, it is apparent that an effective strategy for the reduction of short-term memory demands will have to

meet at least three criteria. First, it must be simple and require little or no operator effort to use. That is, it must not take on the characteristics of a distractor task and compete for the limited working capacity of short-term memory.

Next, it will be most effective if it does not directly require a stimulus transformation on the part of the user. For example, verbal stimuli may be grouped prior to presentation, thus relieving the operator of this burden. Short-term memory mnemonics which require active operator transformation are likely to be heavily influenced by practice and require extensive training (e.g., Reisberg, Rappaport, and O'Shaughnessy, 1984).

Finally, such a strategy must of course lead to a net reduction of the user's mental workload, either by expanding his working span or capacity or by reducing the processing demands of the task itself.

The following section documents some suggestions from the literature on possible strategies for the reduction of short-term memory demands. Some possible approaches have already been suggested in the previous discussion of short-term memory variables.

Grouping

Grouping, or "chunking", is the reorganization of information into meaningful pieces. Short-term memory capacity is more a function of grouping capacity than capacity for bits of information (Miller, 1956). For example, try reading and recalling the following series of letters: LBASLEBA. The same letters, when presented in another fashion become much easier to recall: BASEBALL. One explanation of this phenomenon is that in the second order of presentation, a chunking strategy is very apparent. The letters can be grouped into one English

word.

Conrad, Thomson, and Baddeley (cited by Baddeley, 1982) varied predictability and sequence length of pseudowords and real words in a recall task (Table 4). Not surprisingly, the number of errors per sequence increased systematically with sequence length and dissimilarity to English words. One interpretation of these data is that short, English-looking series of letters may have been easier for subjects to group than long, random strings of letters.

Another helpful grouping strategy is the use of rhythmic grouping. One often uses this technique when repeating a new telephone number. Groupings of three or two are usually best, with a slight time interval between them (Baddeley, 1982).

Finally, Frick (1984) showed that simultaneous presentation of visual information facilitates chunking more readily than does serial presentation.

Hierarchical Organization

Formation of an organizational hierarchy is really a form of grouping on a larger, multidimensional scale. In discussing aviation instrumentation, Loftus and Loftus (1976) illustrated how hierarchical organization principles may be used to subdivide a radar screen into sectors. By assigning distinct decisional rules to each sector of the screen, operators should be better able to attend selectively to visual stimuli. This approach in turn should reduce competing input for the limited capacity of short-term storage.

Hierarchical design principles are also used in the design of instrumentation panel layouts. Figure 14 illustrates part of a possible hierarchical scheme for aircraft instrumentation (Loftus and Loftus, 1976). Such coding schemes may make it

Random Sequences	First Order (Frequency of individual letters the same as in English)	Second order (Pairs of letters equivalent to English)	Third order (Letter triplets equivalent to English)	English words
RCIFODWVIL	TNEOOESHHE	HIRTOCLENO	BETEREASYS	PLANTATION
GKTODKPENF	INOLGGOLVN	DOVEECOFOF	CRAGETTERS	FLASHLIGHT
TZXKHAWCCF	PDOASLOTPP	SESERAICCG	TOWERISBLE	UNCOMMONLY
NGORHQIYWB	AEOCAOIAON	AREDAGORIZ	DEEMEREANY	ALIENATION
BVNJSYZXUA	IRCRENFCTN	CUNSIGOSUR	THERSERCHE	PICKPOCKET

TABLE 4. Words and pseudowords from Conrad, Thomson, and Baddeley, after Baddeley (1982).

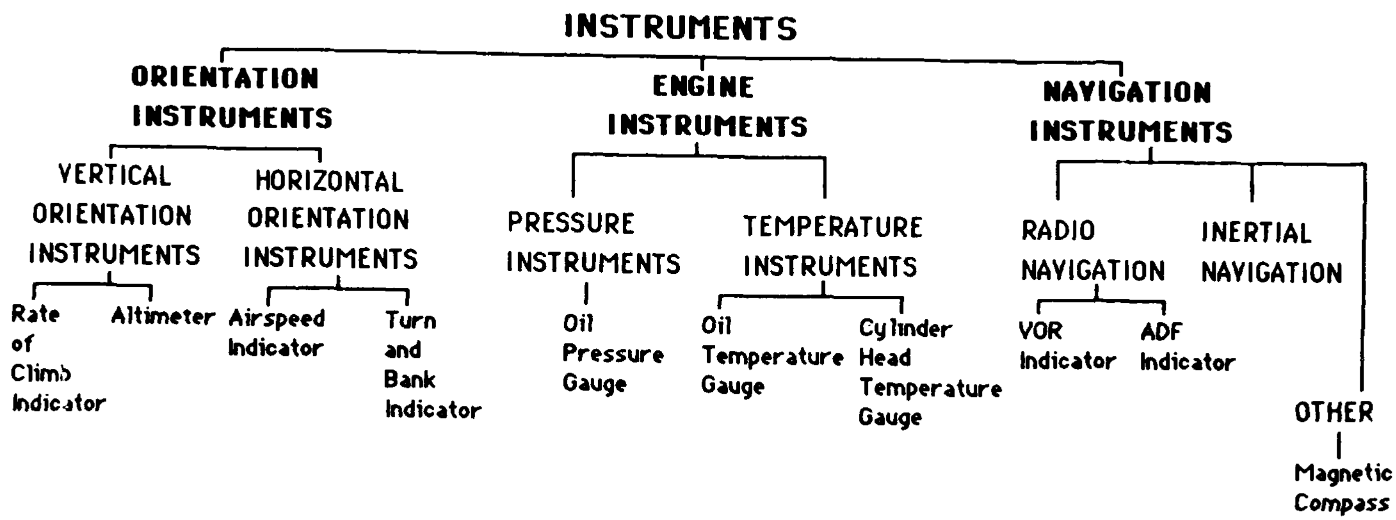


Figure 14. A hierarchical labeling scheme for aircraft instrumentation, after Loftus and Loftus (1976).

necessary for operators to process less information and use their limited short-term memory resources more efficiently.

Distractor Tasks

Two other memory strategies were proposed by Loftus and Loftus (1976) in their discussion of aircraft instrumentation. First, they pointed to the similarity between pilot and ground communication and the Peterson-Brown distractor task. Loftus and Loftus offered the following example:

> The controller may suddenly issue the information that the pilot should change his transponder code to 7227 and contact Seattle Approach Control on radio frequency 119.3. The pilot often thus has to engage in some kind of distractor task (for example, scanning the instruments [or] listening to additional instructions from the controller) before responding [to the controller's instructions] (Loftus and Loftus, 1976, P. 156).

Task structuring so as to minimize or put on hold competing or distracting tasks may be one way to reduce information loss or error generation in short-term memory.

Release from Proactive Inhibition

The second strategy which Loftus and Loftus proposed involves the release from proactive inhibition technique previously discussed. Since pilots' stimuli are often predominately digital in nature, proactive inhibition could accumulate for this stimulus class. The alternation of stimulus class (e.g., using letters for transponder codes) could alleviate this problem. In addition, Loftus and Loftus suggested using a chunking strategy combined with alpha frequency codes to identify their assignment. For example, SEAT, rather than 119.3, could represent the frequency of the Seattle-Tacoma

Control Tower.

Rehearsal

The use of rehearsal to maintain information in the short-term store has been discussed in previous sections. In summary, prevention of rehearsal interference where subvocal, visual, or visuo-spatial rehearsal shows a substantial savings in memory maintenance is one strategy for extending the efficiency of working memory. This of course may be limited to the encoding of stimuli which are semantically or phonetically interpretable, or are relatively simple visual patterns. Alternatively, information processing tasks could be structured so as to limit the need for extended rehearsal.

Dual Storage

Frick (1984) attempted to increase digit span by presenting four digits visually and the remaining digits auditorily. Digit span in this group increased three digits over baseline, although this held only for inverted response conditions (auditory report first, visual report second). Frick suggested that inverted response differences may be due to modality specific interference differences (i.e., reporting a digit in recall verbally will interfere with acoustical memory). Additionally, Frick recommended that, given unequal digit loads in the two stores, the store with the largest digit load should be reported from first. Given the qualifications, these data suggest that the use of nonredundant storage in visual and auditory stores is a viable means of extending immediate digit span.

Adjunctive Rehearsal Mechanisms

Reisberg et al. (1984) adopted a flexible model of working memory to propose the development of an adjunctive rehearsal

mechanism, the finger loop. Viewing working memory slave systems as processing strategies rather than memory stores per se, Reisberg et al. successfully trained subjects in the use of a finger rehearsal strategy for the serial recall of digit sequences. This finger loop is comparable to the Baddeley and Hitch (1974) articulatory loop.

The major results from the study were the following:

1. Subjects learned to increase digit spans by 33% using the finger rehearsal loop;
2. Practice elevated this increase as high as 50%;
3. There was no measurable mental effort (as measured by response latency and rate) for rehearsal of small loads (two digits) with the finger loop; and
4. Both articulatory and finger loops seemed to be tied to motor systems (speech and finger movement, respectively).

Two factors seem to be of additional importance in the implementation of such adjunctive rehearsal systems. First, although highly practiced subjects were able to avoid finger loop interference from concurrent articulatory rehearsal, such was not the case with inexperienced subjects. Secondly, since motor involvement is clearly implicated, providing additional motor feedback (such as a keyboard for the finger loop) could enhance use of the adjunctive rehearsal mechanism.

Redundant Visual Cueing

Simon (1984) investigated the effect of redundant cueing (color and shape) on recall choice reaction time. Subjects were required to discriminate between same or different pairs of stimuli and make a key-press response. A 500-ms interval was used between first stimulus offset and second stimulus onset to assure that short-term retrieval (rather than sensory store retrieval) was being used to make the difference judgements.

Simon's results were mixed, but they did show that the redundant coding group (shape plus color) produced choice reaction times (409 ms) that were significantly faster than the color coding group (540 ms). Mean reaction time for the shape coding group was 456 ms. The possibility of the use of a redundant cueing strategy for increased short-term memory efficiency is an interesting one and merits further empirical investigation.

Automated information Management

The promise of expert system technology as a means of operator aiding could have direct bearing on operator short-term memory limitations. Rasmussen (1981) discussed computer support of operators in process plant fault diagnosis. He concluded that the most important function in such system support would be "to minimize the load upon short term memory" (p.254).

This may be particularly true in cases where what Rasmussen called "decision table" and "hypothesis testing" search strategies are used. In fact, short-term memory constraints may themselves influence the mental search strategy selected by an operator (Rasmussen, 1986).

Rasmussen (1981, 1983) has categorized the behavior of skilled operators as belonging to three groups:

1. Knowledge-Based Behavior;
2. Rule-Based Behavior; and
3. Skill-Based Behavior.

Since it is in rule-based behavior that short-term recall errors are most likely to occur (Rasmussen, 1987), automated aiding of rule-based behavior should yield the greatest relief to an operator's short-term memory resources.

However, the application of computer information management systems is complex and context dependent; automation does not always guarantee reduced mental workload. For example, Goodstein (1981) warned that misapplication of computer controlled information presentation could force operators into a rigid and demanding processing state, "especially with respect to loading of short-term memory" (P. 433).

DISCUSSION

Introduction

A summary glance over the evolution of memory models and their current state leads to some fundamental conclusions. First, most contemporary short-term memory models are built in a linear information processing format. Since the introduction of the digital computer, psychologists and engineers alike have been drawn to this type of model. Pioneering work by Broadbent, Waugh and Norman, and Atkinson and Shiffrin was highly influential in providing the impetus for this movement. Baddeley and Hitch's central processing unit, the central executive, carries the computer model one step further.

Secondly, as models of short-term memory become more complex and seemingly more concrete, there is a temptation to consider the models as more than they are: theoretical constructs. It is pertinent to underscore the fact that short-term memory does not exist, per se. Rather, it is a concept embodied in a large number of models, which attempts to unify a body of data which is both large and varied.

The tentativeness of both the models and the concept are underscored by 1) the failure of any single model to date to account fully for empirical memory phenomena and 2) the continued suggestion by some cognitive scientists (notably,

Craik and Lockhart) that a simple, duplex interpretation of the memory continuum is in error.

Nevertheless, current short-term memory models continue to provide useful theoretical frameworks for cognitive science. In particular, the short-term memory concept may show beneficial application to several facets of the problem of elevated aircrew mental workload.

Short-Term Memory and Expert Systems

Nearly two decades ago, Proctor (1969) concluded that the specification of the man-machine interface was the central problem in the design of command control systems. That problem today may be receiving some answers from the field of artificial intelligence. Kuperman and Wilson (1986) have pointed to the potential use of expert system technology in the management of information in the advanced manned bomber environment. They cited the following possible applications:

1. Threat capability management;
2. Maintanance of nonfixed target inventory;
3. Avionic subsystem management;
4. Integration of onboard data bases and offboard sensors; and
5. Sensor blending and sensor fusion.

Accepting the feasibility of these applications, it then becomes imperative to consider how such an automated system would effect man-in-the-loop performance of aircrews and how the optimization of the man-machine interface may be acheived. In this case, it can be shown that the short-term memory concept is an important element of what Kuperman and Wilson call "a human centered approach to artificial intelligence in the crew station" (p. 44).

Loftus et al. (1979) and Loftus and Loftus (1976) have also

illustrated the relevance of applying the short-term memory concept to the analysis of the aviation environment. Loftus et al. (1979) suggested that current coding in pilot/ground controller communication "has substantial room for improvement in terms of minimizing memory failure" (p. 169).

Thompson (1981) gave this simple example of expert system aiding in commercial air traffic control:

> The ground controller's audio communication with the flight crew may be supplemented by a digital link, so that course/speed/waypoint changes may be entered by the controller into a numeric keyboard supplemented by selected function buttons. The controller's commands would then be transmitted to the aircraft to be displayed on the pilot's navigation CRT as well as heard by him (increasing accuracy and reducing confirmation delays). In the event that the pilot was told to come to 090^{o} (or to reduce speed by 50 knots) and he failed to do so within a reasonable amount of time, he would be automatically alerted to this command (p. 43).

Other issues relevant to such an application include aircrew communication, cockpit annunciator design, and in-flight maintenance checklisting.

Eprath and Young (1981) have illustrated the context specific nature of implementing automatic information management systems. They concluded that in low workload tasks, benefits may accrue from maintaining a high degree of operator involvement in the loop. However, in a complex, high workload system, such benefits are quickly offset by the elevated workload induced in the operator in the loop.

Gaddes and Brady (1981) have established system development guidelines for automated maintenance test programs for

detecting and diagnosing mission avionics faults. According to Gaddes and Brady, the "ideal 'mission failure-free' avionics system" may only be obtained if human performance characteristics (e.g., short-term memory) are accounted for.

The issue of short-term memory is especially germane to highly complex or stressful military aviation scenarios such as low altitude flight and the SRT (Strategic Relocatable Target) mission. Large amounts of information need to be processed by aircrews at greater than ideal rates with many stimuli in direct competition for operator attention. Given the phenomenon of perceptual narrowing in dangerous environments (e.g., Baddeley, 1972), such competition for attention may be especially potent in high-risk military aviation venues. Since the consequences of error are magnified in these scenarios, optimization of the information processing loop should be stressed. Freeing attentional resources by reducing short-term memory demands on aircrew members through automated information management seems at face value a valid approach toward this end.

Short-term Memory and Mental Workload

In light of the previous discussion, it should not be surprising that short-term memory tasks have been incorporated into workload assessment research. These tasks accomplish at least two ends. First, they create a state of mental load which is easily controlled by varying the rate of stimulus presentation, the number of items in a memory set, the duration of a retention interval, etc. Secondly, short-term memory tasks provide their own behavioral indices of mental workload (e.g., recall errors, response latencies, etc.).

Eggemeier, Crabtree, Zingg, Reid, and Shingledecker (1982) used a short-term recall procedure to evaluate the sensitivity of the Subjective Workload Assesment Technique (SWAT). Eggemeier

et al. concluded that SWAT ratings were most sensitive to task difficulty differences in low memory load conditions.

Wickens et al. (1986) cited another example. A short-term memory recognition task (the Sternberg scanning task) was used as a workload diagnostic measure. Although bounded by task-specific limitations, the Sternberg memory search task as a secondary task was capable of revealing the component load sources within the primary task.

Recommendations

The short-term memory concept has held mass appeal for cognitive researchers as well as laymen for over two decades. Because of this, there is a danger that it has become too familiar and is used too freely. It therefore becomes doubly important that a research endeavor attempting to apply this concept to system design first incorporates research into some of the fundamental conceptual relationships involved. In particular, the proposed application of the short-term memory concept in a workload-reducing crew station expert system must be preceded by an initial investigation of the general relationship of short-term memory to mental workload.

While the literature in mental workload is still undecided as to which of a number of diagnostic performance measures is best suited as a general index of workload, short-term memory tasks have played an important role in shaping and validating those measures. For example, recall tasks have been used extensively as both primary and secondary tasks in workload research. Most recently, the Sternberg scanning paradigm has shown strong promise of providing a stable, quantitative description of short-term memory resources, mental workload in general, and the relationship between the two concepts. For these reasons, the Sternberg scanning paradigm was selected for use in the preliminary research conducted in this research effort.

II. MENTAL WORKLOAD LITERATURE REVIEW

INTRODUCTION

Mental workload has received sustained attention in man-machine system research and development over the last decade. Sanders and McCormick (1987) listed the following as possibly beneficial applications of a workload assessment battery:

1. Allocating functions and tasks between humans and machines;
2. Comparing alternative equipment and task designs;
3. Monitoring operators of complex equipment to adapt to task difficulty or allocation of function; and
4. Selecting operators who have higher mental workload capacities for demanding tasks (p. 69).

There is no universally accepted definition of mental workload. However, the construct in its most general form involves two elements: the mental resources of an operator and those resources required by a task. Given this definition, mental workload can be manipulated by changing either operator resources or task demands.

MEASUREMENT OF MENTAL WORKLOAD

A review of the workload assessment literature (Wierwille and Williges,1978) cited 28 techniques that have been used for workload measurements. Most of these techniques can be grouped into one of three categories of workload measures: (1) performance measures, (2) operator activation-level (physiological) studies, and (3) subjective effort ratings.

Performance Measures

Primary task performance. Primary tasks are designed to measure performance on some task-related variable of interest.

Primary task analysis assumes stationarity in the underlying task continuum. For this reason, continuous control tasks are popular choices in primary task paradigms. Some examples of primary task measures include vehicular steering reversals, RMS tracking error, and recall errors. Primary tasks may be used singularly or in combinations of two (dual tasks) or more.

One disadvantage of the primary task methodology is that primary task measures are highly task-specific and consequently it is difficult to compare the workloads imposed among different primary tasks.

Spare mental capacity. The concept of spare mental capacity has been derived from information theory and assumes limited human channel and attention capacity (Rolfe, 1971, cited in Kantowitz and Sorkin, 1983). Two popular spare mental capacity paradigms are time-line analysis and the secondary task paradigm.

Time-line analysis. Time-line analysis uses a task analytic approach in which workload is defined as a function of the time required and the time available to perform the tasks. Sanders and McCormick (1987) cited SWAM (Statistical Workload Assessment Model) as one example of computer-based modeling programs which can accomplish time-line analyses. Stone, Gulick, and Gabriel (1987) recently used time-line analysis to evaluate crew workload in the DC-9. Crew workload (W_I) was defined by the following index:

$$W_I = (T_R / T_A) \times 100 \tag{2}$$

where T_R is the time required to complete an action and T_A is the time available. One disadvantage of time-line analyses is that they do not account for the ability of operators to

timeshare some tasks.

Secondary task measures. The majority of mental workload paradigms use a secondary task format. This paradigm assumes that an operator will divert spare mental capacity from performance of the primary to the secondary task. Greater mental workload will lead to less spare mental capacity and consequently poorer secondary task performance.

Examples of secondary tasks are tapping tasks, arithmetic tasks, choice reaction time tasks, critical tracking tasks, memory search, and time estimation (see, e.g., Casali and Wierwille, 1984; Johannsen, Pfendler, and Stein, 1976; Rolfe, 1976). Casali and Wierwille (1984) found that time estimation was the most sensitive secondary measure for workload on perceptual, mediational, communication, and motor tasks.

Sanders and McCormick (1987) summarized a fundamental problem with the secondary task methodology:

> In order to measure spare resource capacity, the secondary task should tap the same resources as those tapped by the primary task. If one accepts a multiple resource model, then the secondary task should share common modalities (i.e., visual, auditory, speech, motor) and common processing codes. Such a task, however, interferes with the performance of the primary task. Therefore, one cannot say whether the workload measured is imposed by just the primary task or that of the primary as interfered with by the secondary task (Sanders and McCormick, 1987, p. 71).

Because of the difficulty in interpreting secondary task results and controversy surrounding the limited channel model, it may be that a universal secondary task does not exist (Pew, 1979, cited by Kantowitz and Sorkin, 1983). Also, selection of primary and secondary task combinations in recent literature

seems to be unguided by theoretical foundations.

Operator Activation-Level Studies

Operator activation-level studies are based on the assumption that the level of the operator's physiological response to task or system demand depends on his effort. Several different theoretical and experimental studies have demonstrated the relevance of physiological measures in the assessment of mental workload.

Wierwille (1979) compared 14 different physiological measures of aircrew mental workload. According to Wierwille, the use of physiological measures assumes that:

> As operator workload changes, involuntary changes take place in the physiological processes of the human body (body chemistry, nervous system activity, circulatory or respiratory activity, etc.). Consequently, workload may be assessed by the measurement and processing of the appropriate physiological variables (p. 575).

Firth (1973) also suggested that an organism's physiological state reflects its task interactions. He labeled this idea "organic cost."

It is assumed that mental workload greatly influences the activity of the central nervous system (CNS). Therefore, measures of mental workload should reflect some changes in the CNS (Ursin and Ursin, 1979). However, making a distinction between workload activation specific to the perception of the individual operator and the actual workload imposed is not always possible. Therefore, physiological techniques may not accurately reflect the actual amount of the imposed workload, being possibly confounded by the operator's estimate of the workload.

A wide variety of physiological techniques has been evaluated (e.g., Hancock, Meshkati, and Robertson, 1985; Wierwille, 1979). Wierwille (1979) reviewed the 13 most common physiological measurement techniques which are as follows:

1. Heart rate
2. Electrocardiogram
3. Galvanic skin response
4. Muscle tension
5. Electromyelogram
6. Flicker fusion frequency
7. Evoked cortical potentials (P-300)
8. Electroencephalogram
9. Pupillary dilation
10. Eye and eyelid movements
11. Respiration analysis
12. Body fluid analysis
13. Speech pattern analysis

Wierwille (1979) concluded that the most promising physiological measurement techniques seem to be pupil dilation, evoked cortical potentials, and body fluid analysis. However, in all these cases, sophisticated equipment and sometimes intrusive measurement techniques are required to obtain the appropriate data.

He also concluded that no physiological technique alone is likely to provide a valid assessment of mental workload. However, if physiological measurement techniques are combined with behavioral measures, a more adequate description of workload may be obtained.

Most other researchers agree that single physiological measures probably do not provide adequate predictive information to allow assessment of workload. Multiple physiological measures,

used in a combined analysis, usually lead to better assessment and prediction of workload. Techniques such as multiple-regression, correlation, and multivariate analysis (Williges and Wierwille,1979) can be applied to these cases.

Hancock et al. (1985) have reviewed physiological measurement techniques in different perspectives. They placed the various techniques in a two-dimensional space (Figure 15). The abscissa represents a practicality/impracticality scale. This scale is concerned with the question of how practical the measure is under specific conditions. For example, the cost of equipment and operation of the system, ease of the techniques to be used, and the reliability of the measure are all factors of practicality used by Hancock et al.

The ordinate represents the spatial and systemic congruence (SSC) of the measure with respect to the active CNS. Hancock et al. referred to spatial congruence as the actual spatial distance from the CNS. For example, measures of eye/eyelid movement score high on this component of SSC whereas GSR measures score low. Systemic congruence refers to the level of relationship between the physiological function and activity of CNS. Therefore, measures of evoked cortical potentials score high on this component of SSC scale where measures of cardiovascular activity score lower.

Subjective Effort Rating

Subjective opinions can be collected either by rating scales or by questionnaires and/or interviews. A rating scale provides a psychometric technique for ordering opinions in a mathematically consistent manner whereas interview or questionnaire data are not as easily numerically structured.

Subjective rating scales for mental workload estimation have been suggested as the most sensitive and simple measures

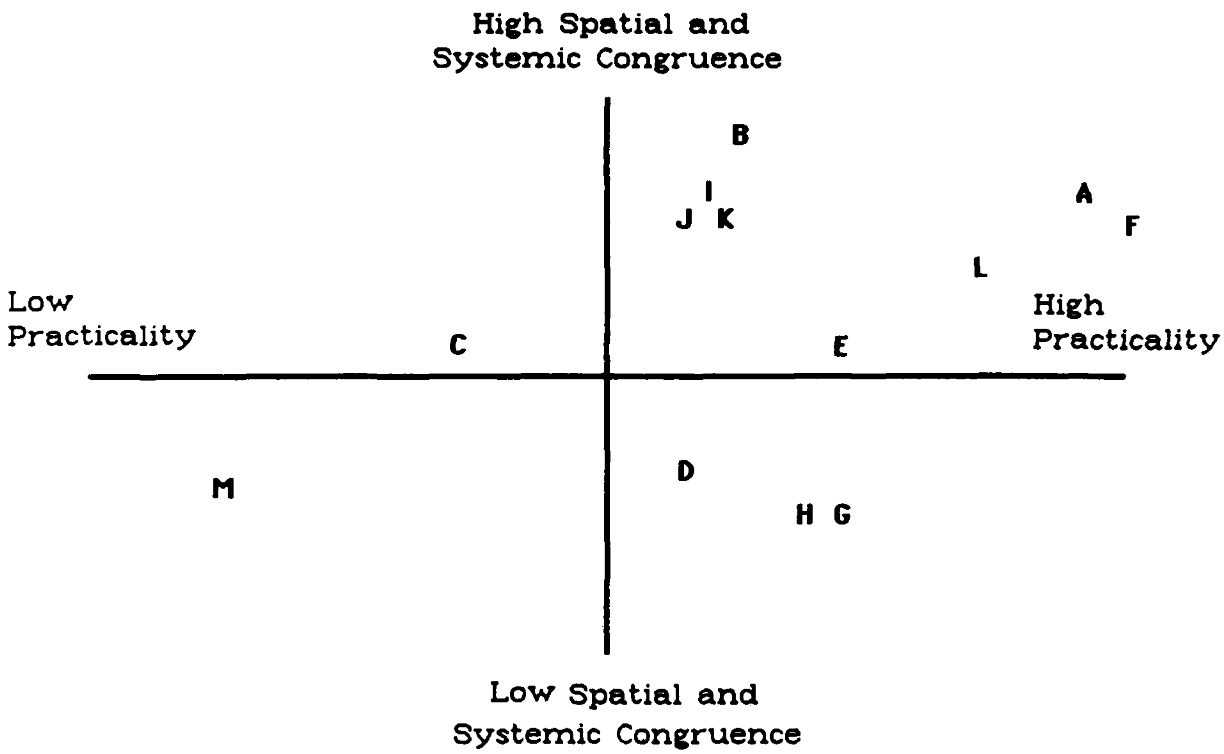


Figure 15. Major physiological workload measures in two-dimensional space, after Hancock, Meshkati, and Robertson (1985). Individual measures are represented as follows: A = Auditory Canal Temperature, B = Event Related Potentials, C = Flicker Fusion Frequency, D = Galvanic Skin Response, E = Electrocardiogram, F = Heart Rate Variability, G = Electromyography, H = Muscle Tension, I = Electroencephalographic Activity, J = Eye and Eyelid Movement, K = Pupillary Dilation, L = Respiration Analysis, M = Body Fluid Analysis

(Skipper, Rieger, and Wierwille, 1986; Wierwille and Casali, 1983; Wierwille and Williges, 1978). Subjective rating techniques require the operator to judge and report the degree of workload experienced during performance of a given task or system function. In addition to their ease of administration, rating scales are widely accepted by the people who are asked to complete them.

Decision-Tree Scales. One type of subjective rating scale is the decision-tree scale. A decision-tree scale is administered in flow chart form. Subjects respond to a series of questions to arrive at a final rating value according to the logic of the decision tree. Skipper et al. (1986) discussed several advantages and disadvantages of decision trees. The advantages include reduced rating variability and provision of "additional guideposts" compared with bipolar scales.

The primary disadvantage of decision-tree scales is that the final rated value has only ordinal properties whereas some rating scales such as SWAT purportedly have interval properties. However, Skipper et al. argued that in most cases the lack of interval property is offset by greater sensitivity to task loading. Thus, the use of decision-tree rating scales may be recommended if they provide greater sensitivity.

The oldest and most well known decision-tree rating scale adapted to mental workload measurement may be the Cooper-Harper scale (Cooper and Harper, 1969). The scale combines a decision tree and a unidimensional 10-point rating scale and is "well suited for workload estimation in manual control systems" (Skipper et al., 1986, p. 586) or psychomotor tasks. Although this scale has been commonly accepted as a standard workload assessment technique in the aviation industry (Wickens, 1984), many researchers have encountered difficulties when attempting to apply this scale to other workload contexts.

Wierwille and Casali (1983) proposed a modified Cooper-Harper rating scale that can be used for perceptual, cognitive, and communication tasks. Wierwille and Casali modified the written descriptions to lend wider applicability to the scale. The descriptions range from (1) very easy, highly desirable, through (5) moderately objectionable difficulty, to (10) impossible. A number of simulator experiments (Casali and Wierwille, 1983; Casali and Wierwille, 1984; Rahimi and Wierwille, 1982) have demonstrated the sensitivity of the modified Cooper-Harper scale to a variety of activities.

SWAT. Another aproach to subjective measurement of mental workload uses the concept as a multidimensional construct. Sheridan and Simpson (1979) proposed three dimensions to define subjective mental workload. Sheridan (1981) characterized these dimensions as "emotion, busy-ness, and problem difficulty" (p. 26). Reid, Shingledecker, and Eggemeier (1981) used these three dimensions to develop the subjective workload assessment technique (SWAT). Eggemeier (1984) described these three dimensions as follows:

> Time load refers to the percentage of time that an operator is busy, and reflects such factors as overlap and interruption among tasks. Mental effort load refers to the degree of attention or concentration required during task performance. Psychological stress load reflects any additional factors that cause operator anxiety or confusion and therefore contribute to subjective mental load (pp. 13-14).

SWAT requires three phases of application. The first phase involves interval scale construction. In this phase 27 possible combinations of each of three dimensions of workload (time load, mental effort load, psychological stress load) are rank ordered. Then, conjoint scaling procedures are used to construct the interval scale (e.g., Nygren, 1982). Depending

on violations of the conjoint axiom tests, anywhere from one scale for all subjects to one scale for each subject may be developed.

Next, an event scoring phase is conducted. Subjects perform ratings of the three workload dimensions for the task being analyzed on a scale from one to three. Thus, for each rating a unique combination of three scores from one to three is collected.

The final phase of SWAT is the conversion of event scores into values on the interval scale(s) developed in the first phase. Detailed procedural guidelines as well as conjoint scaling software are currentlty available in draft form for SWAT users (Armstrong Aerospace Medical Research Laboratory, 1987).

Eggemeier (1984) reported the applicability of SWAT to a number of different tasks and environments. These included laboratory or part-task simulation environments, full mission simulators, and "conditions that are similar to the early stages of system development." Reid (1985) noted that SWAT is most sensitive in moderate to high workload environments.

Eisen and Hendy(1987) have classified six types of tasks in which SWAT is sensitive to workload differences. These are:

1. Tracking (Reid, Shingledecker and Eggmeier, 1981;Vidulich and Tsang,1985);
2. Short-term memory (Eggemeier, Crabtree, Zingg, Reid, and Shingledecker,1982; Eggemeier, Crabtree, and LaPointe, 1983);
3. Spatial transformation (Vidulich and Tsang, 1985);
4. Spatial memory (Eggemeier and Stadler, 1984);
5. Display monitoring (Notestine, 1984); and
6. Multi-faceted tasks of perception, central processing, and motor response (Crabtree, Bateman, and Acton, 1984).

It is important to remember that the three dimensions proposed by Sheridan and Simpson (1979) and used with SWAT were intuitively derived and have not been given a full, empirical validation. In addition, Boyd (1983) found that when the dimensions were independently varied in a task, the ratings of the dimensions were not independent. For example, if time load only was increased in the task, the operators tended to increase their ratings on all three dimensions.

Pro-SWAT. Eggleston and Quinn (1984) have modified SWAT to provide a projective estimate of the operator's mental workload. This modified SWAT, called projective SWAT (Pro-SWAT), is used during the preliminary system design phase. Eggleston and Quinn (1984) described its methodology as follows:

> Pro-SWAT requires task-knowlegeable raters to mentally project themselves into the operation of the defined system, imagine performing the task and then report the magnitude of the workload 'experienced' at selected times (p. 6).

Pro-SWAT is an attractive workload assessment technique for use during the preliminary design of systems because Pro-SWAT requires no mockups, equipment, or simulation.

There are three major areas of concern when applying Pro-SWAT. First, "Its primary limitation is in the ability of the subjects to accurately assess workload based solely on task and equipment descriptions" (Eisen and Hendy, 1987). Second, accurate descriptions of the system design are essential for the subjects to understand the capabilities and limitations of the system. Third, "task-knowledgeable raters" are essential (Eggleston and Quinn, 1984).

In addition to Pro-SWAT, other modifications to the original SWAT can be found in the literature. For example, SWAT 2 may have increased sensitivity for low workload situations (Reid, 1985). Both pro-SWAT and SWAT 2 need further empirical development before they can be applied in a wider variety of task environments.

DISCUSSION

Despite the proliferation of research in this area and all the measurement techniques that have been proposed, there is still no real agreement on a global measure of mental workload. However, researchers do seem to agree that mental workload is a multi-dimensional phenomenon. Therefore, several indices of mental workload will be needed.

In some instances where multiple measures of mental workload have been used together, different measures have provided different results. When this happens, the measures are said to dissociate (McCloy, Derrick, and Wickens, 1983; Yeh and Wickens, 1984). A common finding is that subjective measures dissociate from task performance. Yeh and Wickens (1984) suggested that subjective measures are more sensitive to the number of current tasks being performed, while task performance is more sensitive to the degree of competition for common resources among the various tasks being performed.

Recommendations

A major objective of the current research effort is to substantiate the existence of a dynamic relationship between short-term memory and mental workload. In agreement with the multidimensional nature of mental workload and the data on dissociation of measures, several measures are recommended for inclusion in the research.

Performance measures. It has already been argued in Section 1 of this report that the Sternberg scanning paradigm be used as a secondary task toward this end. A logical candidate task for primary loading is a continuous tracking task. Primary tracking tasks yield the advantages of stationarity, easily manipulated levels of task difficulty, and a degree of face validity when applied to aviation environments.

Physiological measures. The state of physiological indices to date remains one of uncertainty in interpretability. In addition, the likely need for the implementation of a complete battery of such indices to acheive a stable measure of mental workload gives them a low cost-effectiveness. Therefore, the inclusion of physiological measures in the current research effort does not seem warranted and is not recommended.

Subjective measures. The popularity of subjective effort rating scales stems in large part from their ease of application and sensitivity of measurement. However, the selection of a rating tool must be guided by context-specific recommendations. For example, SWAT may be most sensitive in moderate to high workload situations (Reid, 1985) while the Modified Cooper-Harper scale may be most sensitive in low workload situations.

It therefore is most appropriate to incorporate a combination of both of these rating scales. For the purposes of the current research, the use of both the Modified Cooper-Harper scale and SWAT is recommended.

III. INITIAL EXPERIMENT

INTRODUCTION

It was concluded in Section 1 that the proposed application of the short-term memory concept in a workload-reduced crew station environment (i.e., Kuperman and Wilson, 1986) should be preceded by an initial investigation of the general relationship of short-term memory to mental workload. Rationale and recommendations for behavioral indices included in the experiment are found in the preceding literature reviews and their discussions.

Objectives

This experiment was conducted to describe, both qualitatively and quantitatively, the role of short-term memory in mental workload. The objectives were to (1) evaluate the effects of short-term memory loadings (MSET size) and primary task levels on secondary task performance, (2) investigate the associations among subjective workload measures (SWAT and MCH) and task levels, (3) describe the relationships among objective measures of mental workload (choice RT, error percentages, RMS error) and task levels, and (4) explore Sternberg's hypothesis of a linear relationship between short-term memory loading and choice RT.

Short-Term Memory

The Sternberg memory scanning paradigm has been widely used as a secondary task in a variety of experimental settings since its inception by Sternberg (1966). Hence, it is a well validated model of information processing with a large body of existing data. A detailed discussion of this paradigm is found in Section 1. Wickens et al. (1986) listed several guidelines for successful implementation of the Sternberg procedure as a

measure of pilot workload. Those which are immediately germane to this investigation are:

1. Minimize input/output delays (e.g., minimize visual scanning and motor response times);
2. Use short but irregular intertrial intervals;
3. Avoid MSET sizes of 1 or greater than 4;
4. Vary MSET sizes regularly to avoid fatigue and practice effects;
5. Do not use small sample sizes; and
6. Do not impose task overload, since in highly difficult tasks there will be no residual capacity and hence nonmeaningful Sternberg data.

Previous researchers who have reported mixed results using a secondary Sternberg measure have typically violated one or more of these guidelines. For example, Wierwille and Conner (1983) reported data from only six subjects and a single MSET value of five. As Wickens et al. pointed out, large MSETs are often partially forgotten, leading to larger error rates and less interpretable latency data. In addition, Wickens et al. (1983) suggested that visual, rather than auditory secondary Sternberg tasks "... be used in the visual flight environment to guarantee that variations in resource demands be captured" (p.236).

In the current experiment, subject RT in response to various MSET loadings will be used to measure spare mental capacity as an index of operator mental workload. It is expected that increased MSET loading will lead to a decrement in RT performance in concordance with data previously reported using the Sternberg paradigm.

Mental Workload

Subjective assessment. Given the limits of any secondary task sensitivity and the multidimensional nature of the mental

workload construct, it is desirable to include one or more subjective assessment tools in a battery of mental workload measurements. The Cooper-Harper rating scale as modified (MCH) by Wierwille and Casali (1983) should be sensitive to mediational as well as psychomotor loadings. In addition to its reported sensitivity, the ease of administration made this scale an attractive subjective measure for inclusion in this experiment.

SWAT was selected to serve as the second subjective workload measure in this experiment. Reid (1985) suggested that SWAT was most sensitive to moderate to high loadings. Since the MCH has demonstrated sensitivity to low workload situations, SWAT is its logical complement in a multiple instrument battery.

Performance assessment. The primary tasks chosen for use in this experiment with the secondary Sternberg task are compensatory tracking and visual choice RT. The use of primary psychomotor tasks has a large precedence in the secondary task literature, especially tracking tasks, and their advantages in this regard have already been discussed. However, their frequent use when the secondary task of choice is primarily perceptual or mediational in nature is somewhat inconsistent with multiple resource taxonomies (e.g., Berliner, Angell, and Shearer, 1964) which suggest that these tasks draw from different (although undoubtedly related) resource reservoirs. Given the rather large body of data which has recently supported the basic tennets of the multiple resources concept, it is reasonable to hypothesize that workload variation in a task with more mediational loading (e.g., choice RT) may have a more potent effect on secondary Sternberg task performance, which is highly mediational in nature. Therefore, a visual choice RT task was chosen for inclusion in this experiment as a "dual primary" task.

METHOD

Subjects

Eighteen male students (18 to 29 years of age) participated in this experiment. Subjects were recruited through an advertisement in the school newspaper and were paid $30 each for participation in the experiment. All subjects were required to pass a three-part screening procedure.

Subjects were first screened for visual acuity and phoria with a Bausch and Lomb Master Ortho-rater. Criteria for this test were a normal or corrected Snellen acuity of at least 20/25 and phoria scores within the 88th percentile. Visual phoria is a measure of the tendency of the eyes to turn away from each other in the absence of a stimulus to fusion.

Subjects were next screened for contrast sensitivity with the Vistech Vision Contrast Test System. Contrast sensitivity is normally correlated with Snellen acuity and no subjects who passed the Ortho-rater test failed the contrast sensitivity test.

The last screening test was for tracking performance and basic visual-motor coordination. The test was abstracted from the actual experimental tracking procedure. Based on pilot subject performance, the tracking criterion was set at 20 or fewer RMS error scores (formula 3) which exceeded three pixels during the 40-second screening trial. RMS errors were based on averages of 22 samples/s. The average RMS error for the 18 subjects accepted into the study was 2.6 pixels, while the RMS error for the five subjects who failed the test ranged from 3.6 to 7.0 pixels.

Apparatus, Stimuli, and Experimental Environment

A Texas Instruments Business-Pro computer was used to drive the monitor, generate auditory tones, and to collect data (i.e., tracking performance, subjective workload ratings, and RT data). A 33-cm (diagonal) color Texas Instruments Business-Pro monitor was used for the visual stimulus display. Appendix A contains representations of the screen stimuli used.

The center of the screen was positioned approximately 50 cm from the subject's eyes and 105 cm from floor height. The distance from the table surface to the screen center was 30 cm. All target stimuli were green, with a red screen border and a black background. Space-averaged luminance values for stimuli displayed on the monitor were between 9.1 cd/m^2 for the green stimuli to 2.7 cd/m^2 for the red border. The luminance of the black background was 0.1 cd/m^2. Average luminance of the white wall behind the screen was 1.6 cd/m^2. An incandescent ceiling lamp provided approximately 4.2 lux of illumination measured at the center of the monitor surface.

Auditory stimuli included a 75-ms, 400-Hz tone, a 75-ms, 1000-Hz tone, and a pair of 75-ms, 1000-Hz tones spaced by 50 ms.

Input devices included an isometric (force) joystick (Measurement Systems model 462) and a two-button keypad to collect two-choice RT input. Turbo Pascal software (Borland International, Version 3.0, 1985) was developed for screen formatting, data collection and storage, and RT measurement. The internal MS-DOS clock was used for program control and RT measurement (10-ms resolution).

Skipper et al. (1986) used a computerized version of the MCH rating scale and found ratings comparable to the standard MCH, with some minor variations. Based on these results and a short

pilot study, it was decided to construct a new display format for this experiment. Since subjects were instructed to make their MCH ratings based on the written descriptions rather than the numbers, the numbers were removed form the MCH rating screens (pilot data indicated that in light of these instructions, subjects were confused by the inclusion of numbers on the MCH scale screens). Sample computer MCH and SWAT rating screens are shown in Appendix B and Appendix C, and the original MCH rating scale is shown in Apppendix D.

Design

Primary task (two levels). A dual primary task procedure was employed using a two-axis compensatory tracking task and a two-choice visual discrimination task (Figure 16). At the first level, subjects performed only the tracking task which involved maintaining cursor (crosshair) position in the center of a target (box). At the second (dual) primary task level, subjects were required to identify a "pop-up" missile symbol. Subjects were instructed to report, as quickly as possible but without making a mistake, whether the missile was solid or hollow by pressing the suitable key on the keypad.

Tracking difficulty (three levels). In the compensatory tracking task, the cursor (crosshair) position was driven from the target by disturbances constructed from two combined sine waves. Three levels of tracking disturbance were used to represent "low," "medium," and "high" primary task levels. The three disturbance levels were differentiated by combinations of frequency ratios and sampling rates.

Order (six levels). Level of tracking difficulty was blocked across the three days of data collection. Consequently, there were six possible orders (Table 5) in which the subjects could receive the tracking levels. Subjects were randomly assigned to one of the six levels, with a total of three subjects at

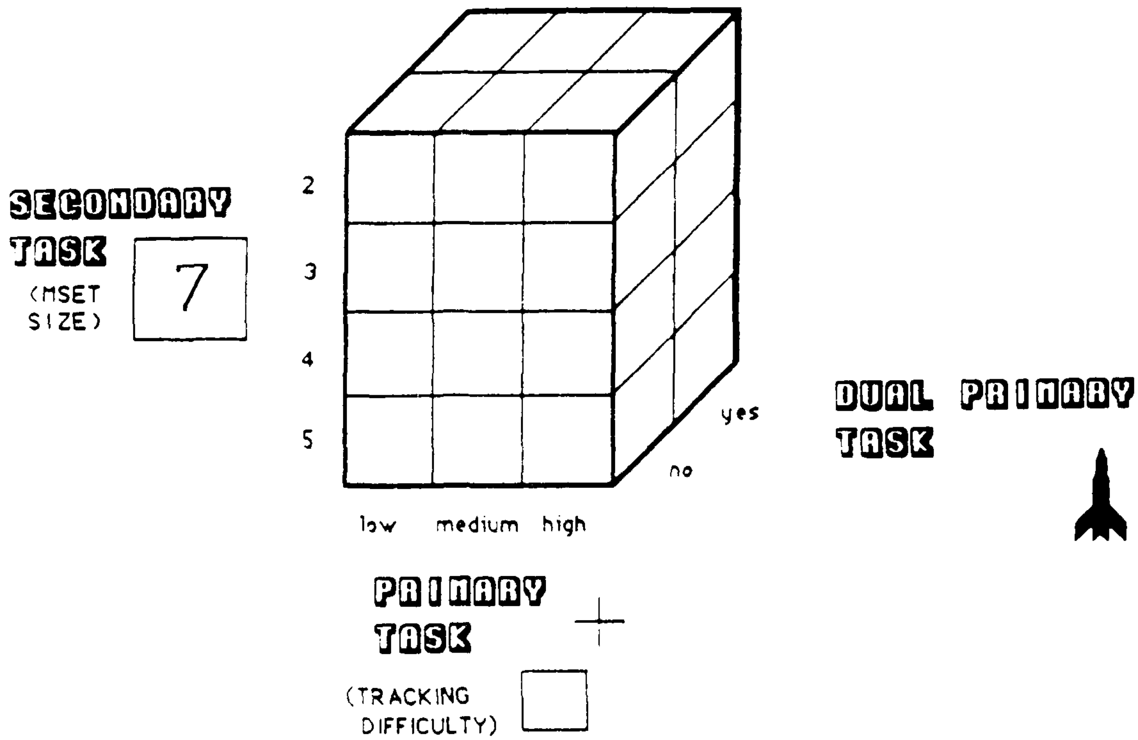


Figure 16. Three-dimensional representation of independent variables

TABLE 5. Sequence of Tracking Difficulty for Order Groups.

	Day		
	1	2	3
Order			
1	Low	Medium	High
2	Medium	Low	High
3	High	Low	Medium
4	Low	High	Medium
5	High	Medium	Low
6	Medium	High	Low

each level.

Secondary task (four levels). A visual Sternberg memory scanning task was used with MSET sizes of 2, 3, 4, and 5 elements with digits 0-9. Stimuli were serially presented from a random, nonrepeating set. The probability of probe inclusion in either positive or negative MSETs was 50%. The rate of MSET presentation was 1 digit/s. The probe followed the last MSET stimuls after an interval of 0.5 to 1.5 s.

Replications (five levels). During the data collection sessions, subjects were given five blocks of trials, each block of trials containing a random presentation of all eight primary task and secondary task combinations. These replications were performed to obtain a more stable measure of performance within each experimental cell.

Trial segments. Each individual 40-second trial was structured into four 10-second segments (Figure 17). The tracking task was performed across all segments of all trials. The dual

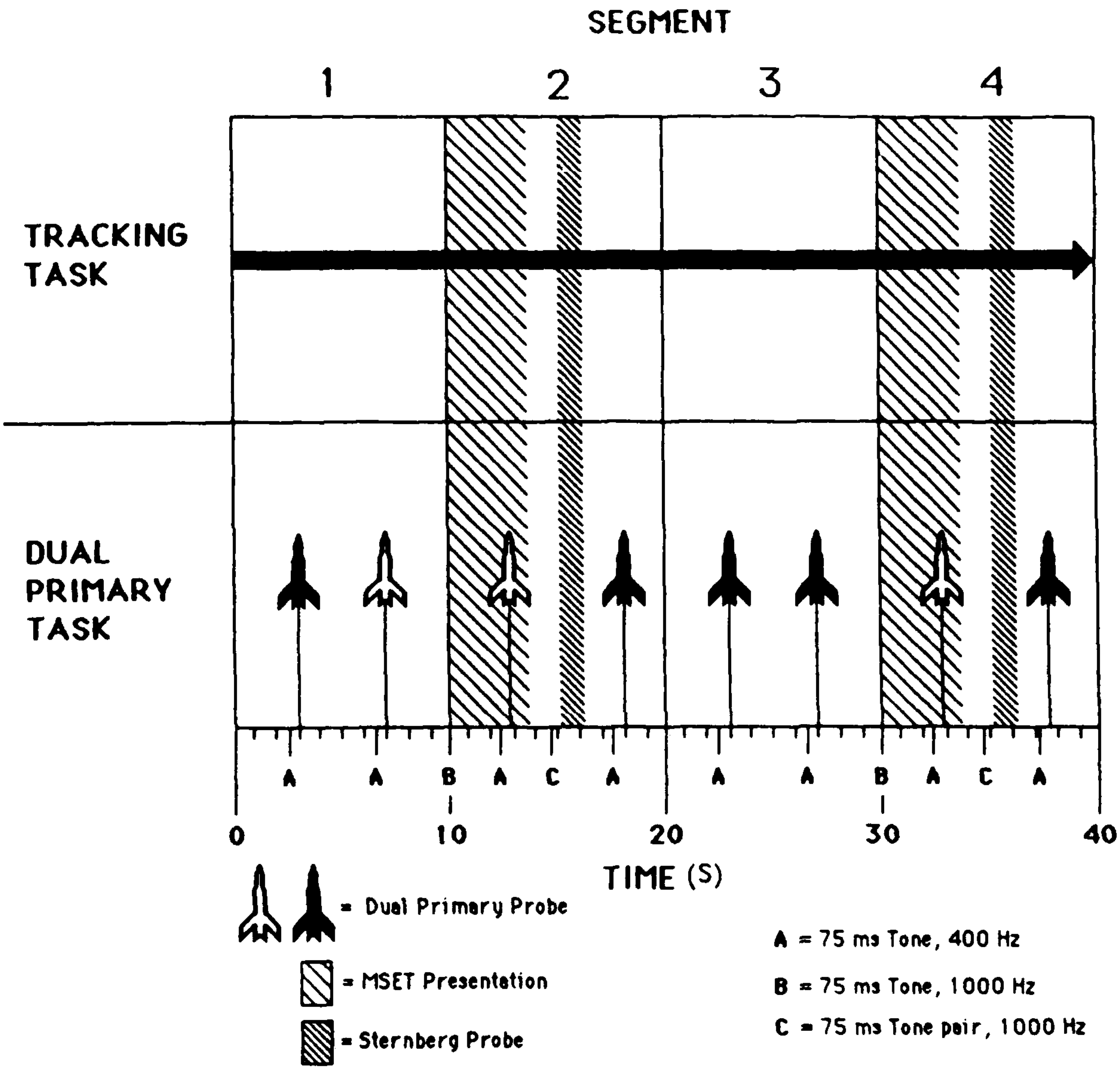


Figure 17. Representative trial segments for MSET size = 3, with dual primary task.

primary task (missile identification) was presented in half the trials with two probes appearing randomly within each segment. The secondary task was presented in all trials during segments 2 and 4 with one MSET per segment. Therefore, any given trial could contain up to 8 dual primary probes and 2 MSETs.

RMS error. A measure of the quality of control in the compensatory tracking task was recorded for each second of each 40-second trial. Each one-second measure was based on a 22-sample average. Poulton (1974) endorsed the use of RMS error as "the measure of overall adequacy of tracking" (p. 38). RMS error (in pixels) for each trial was calculated as:

$$\text{RMS error} = \frac{(\Sigma\ [(X - X_t)^2 + (Y - Y_t)^2\ ])^{1/2}}{N} \qquad (3)$$

where (X,Y) is the cursor position, (X_t, Y_t) is the target position, and N is the number of samples per trial. Low RMS error indicates high quality of control.

Reaction time. The time (in ms) from stimulus onset to completion of motor choice response was defined as RT. These data were collected for both the dual primary and secondary tasks.

Error percentage. The percentage of trials in which the incorrect response was selected was collected for both the dual primary and secondary tasks.

Subjective ratings. Following each trial, subjects rated the workload associated with that trial using computer-presented versions of the MCH and SWAT rating scales. The order of presentation of the two scales was randomly varied.

Procedure

Each subject participated in the experiment over a four-day period. The total average time of participation, including screening and practice, was 255 minutes.

On the first day, subjects received instructions and 12 practice trials on a representative range of primary, dual primary, and secondary tasks. Subjects also received instructions for both subjective ratings scales but did not practice using the computerized rating scales. During the remainder of the first day, subjects completed the SWAT scale development (card sorting) procedure. All instructions (Appendix E) were read aloud by the experimenter as the subject read along. Instructions for the MCH rating scale were adapted from Casali (1982).

Data collection was conducted on the remaining three days. Each day, subjects began by reviewing abbreviated instructions and completing six practice trials. Subjects then completed five 40-second trials for each of the eight primary and secondary task combinations, with each trial immediately followed by the computerized rating procedures. After 24 experimental trials, subjects were given a five-minute rest break.

Subjects operated the joystick with their preferred hand (17 of 18 subjects were right handed) and rested their nonpreferred hand on the keypad.

RESULTS

Subjective Ratings

SWAT scale values were calculated using the additive polynomial

model incorporated in the SWAT conjoint analysis software (Armstrong Aerospace Medical Research Laboratory, 1987). Scales for each SWAT prototype (time, stress, and effort) were developed. Seven of 18 subjects repeated the SWAT scale development procedure (card sort) due to unacceptably high numbers of conjoint scaling axiom violations. Table 6 shows the Kendall's measures of concordance, the number of axiom violations, and the number of subjects associated with each prototype scale developed. The lowest Kendall's W occurred in the effort group (.8766) while the effort and stress groups had the largest total number of axiom violations (32).

The degree of associations among mean SWAT ratings, median MCH ratings, and mean secondary task reaction times (RTs) were calculated with the Spearman Rank Order Correlation procedure. The SWAT and MCH ratings were highly correlated (Rho = .906, $p < .0001$). Although SWAT ratings were significantly correlated with secondary task RTs (Rho = .609, $p = .016$), MCH ratings were not (Rho = .365, $p = .0790$). The mean SWAT and median MCH ratings are presented in Table 7, sorted by mean secondary task RTs, and in Table 8, sorted by MSET size. As seen in Table 8, mean SWAT ratings and median MCH ratings

TABLE 6. SWAT Scale Development Data for the Three Prototype Groups Used.

		Prototype	
	Time	Effort	Stress
Subjects	7	7	4
Kendall's W	.8804	.8766	.9268
Axiom Violations			
Independence	0	14	8
Double Cancellation	1	0	0
Joint Independence	13	18	24

TABLE 7

Subjective Ratings and Mean Secondary Task Reaction Time as a Function of Task Levels*

Mean SWAT	Median MCH	Primary Task Level	Tracking Difficulty	MSET Size	Mean RT
8.14	1	single	low	3	922.20
8.18	1	single	low	2	881.39
9.54	1	single	medium	2	866.89
12.54	2	single	medium	3	905.67
14.43	2	single	low	4	950.29
16.95	3	single	medium	4	951.11
18.99	2	single	medium	5	1017.44
20.50	2	single	low	5	985.17
25.48	2	single	high	2	922.50
27.65	2	single	high	3	957.28
31.12	2	single	high	4	989.66
34.64	2	single	high	5	1014.44
45.81	3	dual	low	2	883.02
49.83	3	dual	low	3	939.55
51.65	3	dual	medium	3	946.93
52.52	3	dual	low	4	945.78
53.47	3	dual	medium	2	885.64
56.46	3	dual	medium	4	1002.22
59.45	3	dual	high	2	954.58
61.11	3	dual	low	5	1006.00
63.60	3	dual	high	3	1063.85
64.09	3	dual	medium	5	1004.11
68.43	3	dual	high	4	1026.89
72.10	3	dual	high	5	1063.72

* Values are sorted in ascending order by mean SWAT rating

TABLE 8

Subjective Ratings and Mean Secondary Task Reaction Time as a Function of Task Levels*

Mean SWAT	Median MCH	Primary Task Level	Tracking Difficulty	MSET Size	Mean RT
8.18	1	single	low	2	881.39
9.54	1	single	medium	2	866.89
25.48	2	single	high	2	922.50
45.81	3	dual	low	2	883.02
53.47	3	dual	medium	2	885.64
59.45	3	dual	high	2	954.58
8.14	1	single	low	3	922.20
12.54	2	single	medium	3	905.67
27.65	2	single	high	3	957.28
49.83	3	dual	low	3	939.55
51.65	3	dual	medium	3	946.93
63.60	3	dual	high	3	1063.85
14.43	2	single	low	4	950.29
16.95	3	single	medium	4	951.11
31.12	2	single	high	4	989.66
52.52	3	dual	low	4	945.78
56.46	3	dual	medium	4	1002.22
68.43	3	dual	high	4	1026.89
18.99	2	single	medium	5	1017.44
20.50	2	single	low	5	985.17
34.64	2	single	high	5	1014.44
61.11	3	dual	low	5	1006.00
64.09	3	dual	medium	5	1004.11
72.10	3	dual	high	5	1063.72

* Values are sorted in ascending order by MSET size

increased roughly as a function of MSET size, primary task level, and tracking difficulty.

Secondary Task Performance

A five-way analysis of variance (ANOVA) was performed on the data to analyze the Sternberg recognition memory performance in terms of RT and error percentage. A 2 x 2 x 3 x 4 x 6 ANOVA (Primary task level by MSET type (positive or negative) by Tracking difficulty by MSET size by Order) was performed on secondary task RT (Table 9). There were five significant main effects: Primary task level (p = .0043), MSET size (p < .0001), Order (p = .0021), Tracking difficulty (p = .0004), and MSET type (p = .0486). Reaction times were significantly greater in the dual primary task condition than in the single primary task condition. Post hoc Newman-Keuls Tests were performed on both MSET size and Order to determine which means were statistically different. The mean RT for MSET sizes 3 and 4 were not significantly different, with MSET size 2 significantly less than and MSET size 5 significantly greater than MSET sizes 3 and 4 (Table 10). The only significant difference among the Order means occurred at Order 3 (p = .0021), having a larger RT than the other orders (Table 11). The average RTs are plotted for the significant main effects of Primary task level (Figure 18), MSET size (Figure 19), and Order (Figure 20).

There was a significant two-way interaction between Tracking difficulty and MSET type (p =. 0393, Figure 21). In general, as tracking difficulty increased, RTs increased, negative MSET RTs more so than positive MSET RTs Post hoc Simple-Effect F-Tests revealed that the MSET Type means were significantly different only at the highest level of Tracking difficulty (p < .01, Table 12).

An identical ANOVA was performed for secondary task error percentage (Table 13). Only one main effect, Primary task

TABLE 9

ANOVA Summary Table for Secondary Task Reaction Time

Source	df	SS	MS	F	p
Order	5	190780.6020	38156.1204	7.53	.0021
Subject (Order)	12	60816.0136	5068.0011		
Tracking	2	6514.5232	3257.2616	11.11	.0004
Tracking * Order	10	6200.4647	620.0465	2.11	.0646
Tracking * Subject (Order)	24	7037.8010	293.2417		
Primary Task	1	2075.5792	2075.5792	12.32	.0043
Primary Task * Order	5	187.4076	37.4815	0.22	.9458
Primary Task * Subjects (Order)	12	2021.3135	168.4428		
MSET	3	15319.4828	5106.4943	22.02	.0001
MSET * Order	15	1594.7756	106.3184	0.46	.9464
MSET * Subjects (Order)	36	8348.9299	231.9147		
Type	1	1710.4391	1710.4391	4.81	.0486
Type * Order	5	1180.5618	236.1124	0.66	.6574
Type * Subject (Order)	12	4262.8448	355.2371		
Tracking * Primary Task	2	516.5382	258.2691	1.14	.3364
Tracking * Primary Task * Order	10	1074.7489	107.4749	0.47	.8901
Tracking * PT * Subj (Order)	24	5434.5664	226.4403		
Tracking * MSET	6	867.8837	144.6473	1.01	.4280
Tracking * MSET * Order	30	5219.6110	173.9870	1.21	.2519
Tracking * MSET * Subj (Order)	72	10348.1209	143.7239		
MSET * Primary Task	3	617.7052	205.9017	1.26	.3026
MSET * Primary Task * Order	15	1027.7259	68.5151	0.42	.9631
MSET * PT * Subj (Order)	36	5881.7670	163.3824		
Tracking * Type	2	1609.6652	804.8326	3.72	.0393
Tracking * Type * Order	10	2874.2791	287.4279	1.33	.2722
Tracking * Type * Subject (Order)	24	5199.0148	216.6256		
Type * Primary Task	1	263.6624	263.6624	1.07	.3217
Type * Primary Task * Order	5	838.1938	167.6388	0.68	.6477
Type * Primary Task * Subj (Order)	12	2961.9130	246.8261		
MSET * Type	3	1238.3060	412.7687	2.57	.0694
MSET * Type * Order	15	3206.1124	213.7408	1.33	.2349
MSET * Type * Subjects (Order)	36	5781.2442	160.5901		

(ANOVA summary table continued on next page)

TABLE 9 (continued)

ANOVA Summary Table for Secondary Task Reaction Time

Source	df	SS	MS	F	p
Tracking * MSET * Primary Task	6	754.5085	125.7510	0.71	.6401
Tracking * MSET * PT * Order	30	5193.3161	173.1105	0.98	.5067
Trk* MSET * PT * Subj (Order)	72	12694.3536	176.3105		
Tracking * Type * Primary Task	2	100.2903	50.1452	0.30	.7412
Tracking* Type * PT * Order	10	1830.4519	183.0452	1.11	.3964
Trk * Type * PT * Subj (Order)	24	3969.2236	165.3843		
Tracking * MSET * Type	6	1365.9083	227.6514	1.69	.1362
Tracking * MSET * Type * Order	30	3419.9336	113.9978	0.85	.6896
Trk * MSET * Type * Subj (Order)	72	9708.7435	134.8437		
MSET * Type * Primary Task	3	780.4965	260.1655	2.00	.1316
MSET * Type * PT * Order	15	2455.2963	163.6864	1.26	.2778
MSET * Type * PT * Subj (Order)	36	4686.0480	130.1680		
Tracking * MSET * Type * PT	6	876.1019	146.0170	0.81	.5629
Trk * MSET * Type * PT * Order	30	4867.1950	162.2398	0.90	.6112
Trk * MST * Typ * PT * Sub (Order)	72	12923.7127	179.4960		

TABLE 10

Results of Newman-Keuls Test on MSET Size (Mean Secondary Task ReactionTime (ms))*

MSET size:	2	3	4	5
Mean Value:	900.77	956.91	976.00	1017.90

*means with a common line do not differ significantly at $p < .05$.

TABLE 11

Results of Newman-Keuls Test on Order (Mean Secondary Task Reaction Time (ms))*

Order Group:	5	1	2	4	6	3
Mean Value:	811.15	834.18	936.04	939.05	992.87	1264.07

*means with a common line do not differ significantly at p <.05.

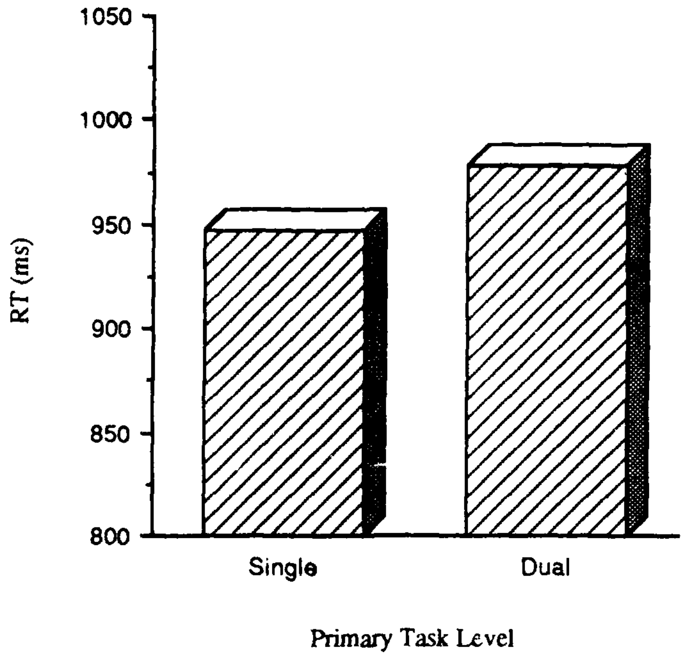


FIGURE 18. Secondary Task Reaction Time as a Function of Primary Task Level.

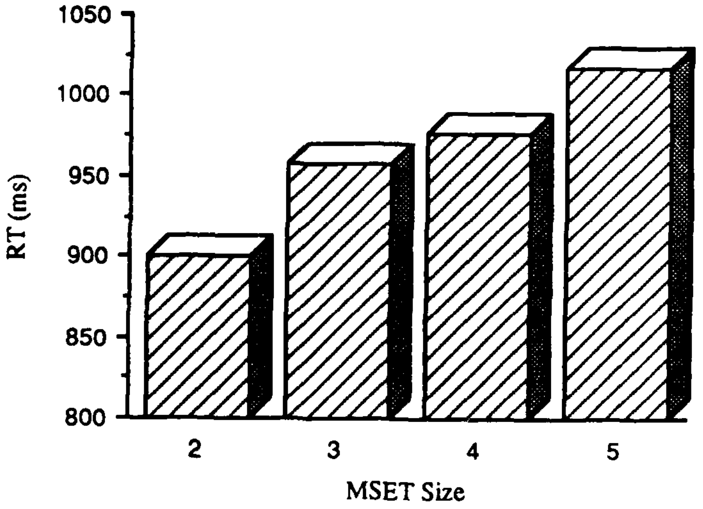


Figure19. Secondary Task Reaction Time as a Function of MSET Size.

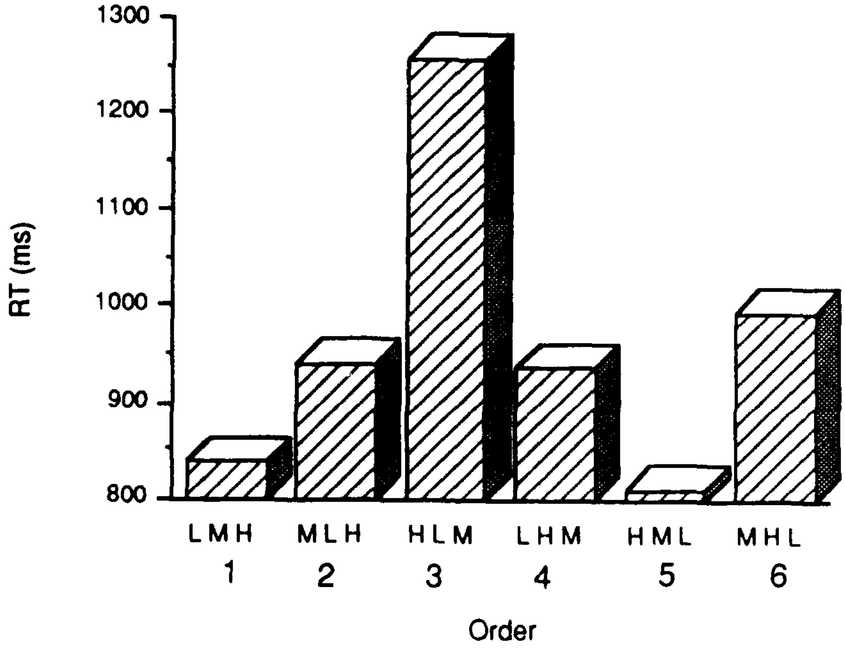


Figure 20. Secondary Task Reaction Time as a Function of Order.

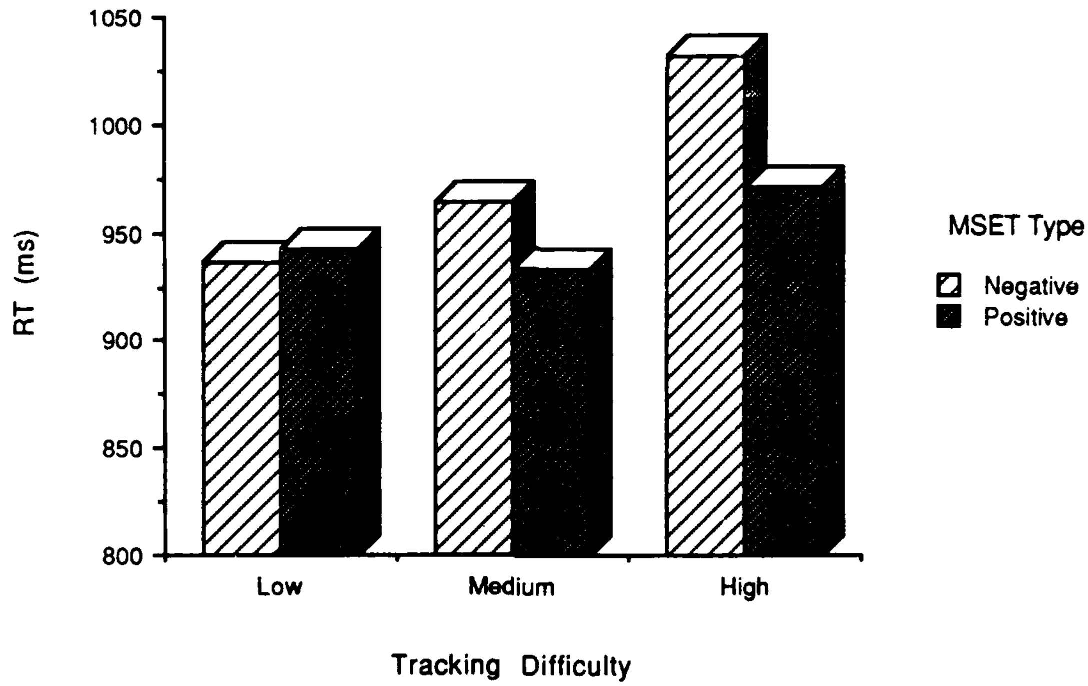


FIGURE 21. Secondary Task Reaction Time as a Function of Tracking Difficulty and MSET Type.

TABLE 12

Results of Simple-Effect F-Tests on MSET Type for Each Level of Tracking Difficulty (Secondary Task Reaction Time (ms))

Tracking Difficulty	$MS_{MSET\ Type}$	F	p
Low	30.0490	.14	> .25
Medium	674.6337	3.11	< .10
High	2615.4217	12.07	< .01

TABLE 13

ANOVA Summary Table for Secondary Task Error Percentage

Source	df	SS	MS	F	p
Order	5	2900.3825	580.0765	2.85	.0640
Subjects (Order)	12	2446.4037	203.8670		
Tracking	2	361.9954	180.9977	1.24	.3078
Tracking * Order	10	510.7846	51.0785	0.35	.9567
Tracking * Subjects (Order)	24	3508.1605	146.1734		
Primary Task	1	2662.4979	2662.4979	16.09	.0017
Primary Task * Order	5	452.5320	90.5064	0.55	.7379
Primary Task * Subjects (Order)	12	1985.3339	165.4445		
MSET	3	115.0547	38.3516	0.26	.8526
MSET * Order	15	538.4178	35.8945	0.24	.9974
MSET * Subjects (Order)	36	5279.7395	146.6594		
Type	1	60.4476	60.4476	0.19	.6705
Type * Order	5	246.5108	49.3022	0.16	.9743
Type * Subjects (Order)	12	3813.9261	317.8271		
Tracking * Primary Task	2	202.0057	101.0029	0.82	.4541
Tracking * Primary Task * Order	10	1158.6431	115.8643	0.94	.5191
Tracking * PT * Subj (Order)	24	2970.7097	123.7796		
Tracking * MSET	6	272.6601	45.4434	0.42	.8648
Tracking * MSET * Order	30	3873.2925	129.1098	1.18	.2766
Tracking * MSET * Subj (Order)	72	7856.1612	109.1134		
MSET * Primary Task	3	224.3016	74.7672	0.58	.6331
MSET * Primary Task * Order	15	1301.8098	86.7873	0.67	.7944
MSET * PT * Subj (Order)	36	4655.7353	129.3260		
Tracking * Type	2	80.2145	40.1073	0.36	.7021
Tracking * Type * Order	10	719.3424	71.9342	0.64	.7624
Tracking * Type * Subjects (Order)	24	2681.7954	111.7415		
Type * Primary Task	1	131.7739	131.7739	0.73	.4110
Type * Primary Task * Order	5	298.9230	59.7846	0.33	.8858
Type * Primary Task * Subj (Order)	12	2179.2465	181.6039		
MSET * Type	3	407.9805	135.9935	1.80	.1647
MSET * Type * Order	15	1680.9154	112.0610	1.48	.1638
MSET * Type * Subjects (Order)	36	2720.6066	75.5724		

(ANOVA summary table continued on next page)

TABLE 13 (continued)

ANOVA Summary Table for Secondary Task Error Percentage

Source	df	SS	MS	F	p
Tracking * MSET * Primary Task	6	167.7707	27.9618	0.31	.9275
Tracking * MSET * PT * Order	30	3681.5037	122.7168	1.38	.1345
Trk * MSET * PT * Subj (Order)	72	6404.5646	88.95		
Tracking * Type * Primary Task	2	14.7436	7.3718	0.05	.9507
Tracking * Type * PT * Order	10	1728.2422	172.8242	1.19	.3463
Trk * Type * PT * Subj (Order)	24	3492.2946	145.5123		
Tracking * MSET * Type	6	393.9181	65.6530	0.63	.7052
Tracking * MSET * Type * Order	30	2688.8799	89.6293	0.86	.6690
Trk * MSET * Type * Subj (Order)	72	7495.4899	104.1040		
MSET * Type * Primary Task	3	105.4878	35.1626	0.41	.7459
MSET * Type * PT * Order	15	1359.2370	90.6158	1.06	.4235
MSET * Type * PT * Subj (Order)	36	3077.9881	85.4997		
Tracking * MSET * Type * PT	6	525.0025	87.5004	1.04	.4048
Trk * MSET * Type * PT * Order	30	3261.0670	108.7022	1.30	.1851
Trk * MST * Type * PT * Subj (Ord)	72	6038.4697	83.8676		

level, was statistically significant ($p = .0017$). Approximately twice as many secondary task errors were committed in the dual primary task condition than in the single primary task condition, as illustrated in Figure 22. The average secondary task error percentage was 5.1 %.

A least-squares linear regression procedure was performed on the secondary task RT data. The data were collapsed across subjects and replications to obtain a stable, average performance measure at each level of tracking difficulty. Figure 23 shows the best-fit linear functions for MSET size by primary task level and MSET type at the low tracking difficulty level, while figures 24 and 25 show the data at the medium and high difficulty levels. These figures reflect the two-way interaction of MSET type and Tracking difficulty on secondary task RT. Slopes of the respective lines were 35.02 ms, 45.21 ms, and 29.77 ms per MSET digit, with the lowest slope value at the highest level of tracking difficulty. Values of R^2 were .64, .92, and .66, with the best fit being at the medium tracking difficulty level.

Primary Task Performance

Three separate ANOVAs were performed to assess the degree of task intrusion on primary and dual primary task performance. Task intrusion on tracking performance was assessed in a 2 x 3 x 4 x 6 ANOVA (Primary task level by Tracking difficulty by MSET size by Order). Table 14 shows the summary table for this ANOVA. Three significant main effects were found in this ANOVA: Tracking difficulty ($p < .001$), Primary task level ($p < .001$), and MSET size ($p = .0181$). These three main effects are shown in Figures 26, 27, and 28.

RMS tracking error was significantly higher in the dual primary task condition than in the single primary task condition. A Newman-Keuls Test on Tracking difficulty means

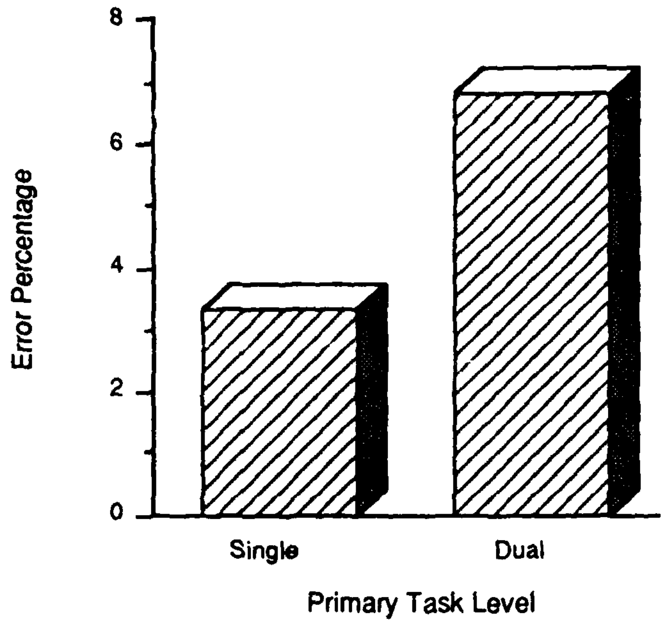
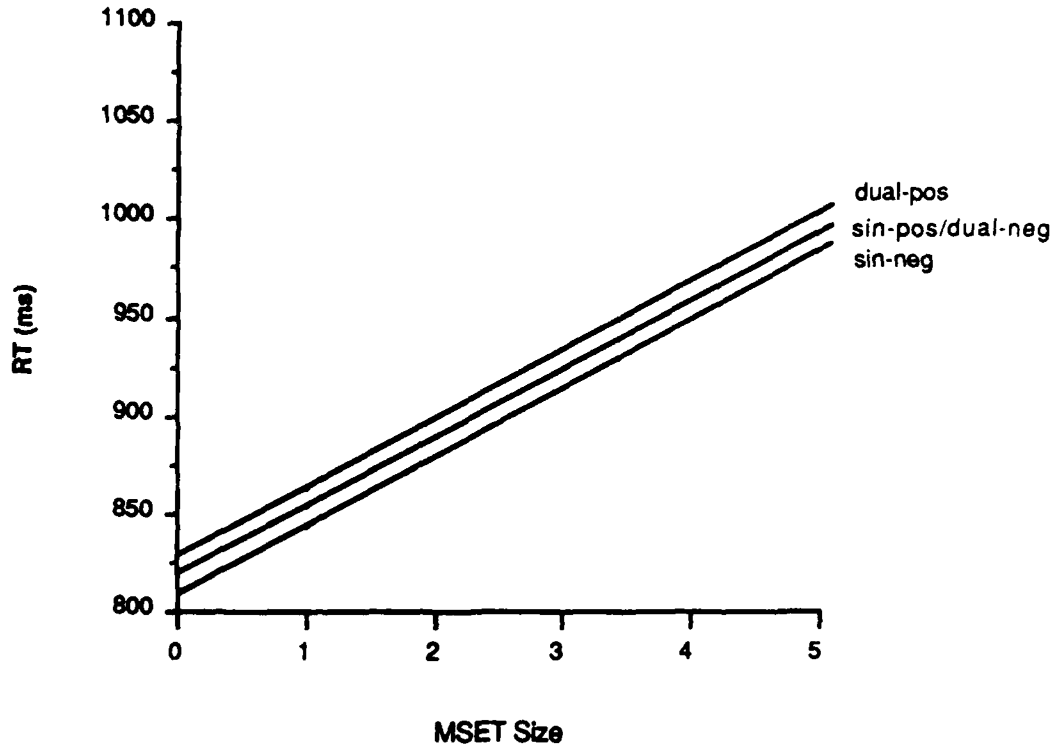


Figure 22. Secondary Task Error Percentage as a Function of Primary Task Level.

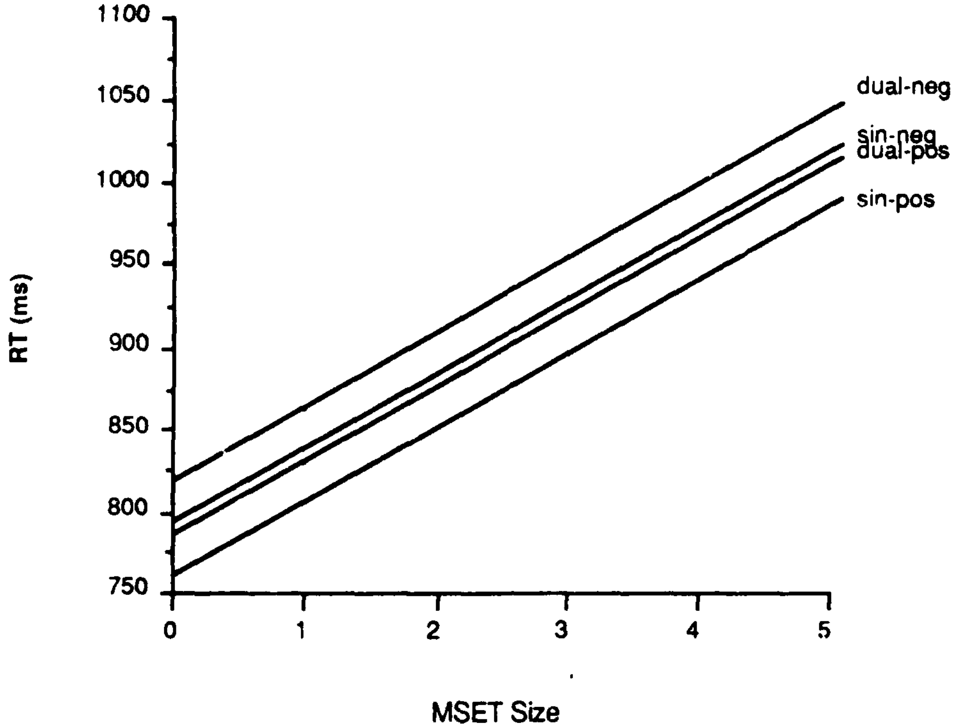


$$RT = 809.06 + 35.02\ MSET + 9.93\ PT + 9.97\ Type$$

$$R^2 = .64^*$$

Figure 23. Least Squares Linear Regression Across MSET for Tracking Level = Low (categorical variables were included to represent when a dual primary task was present (PT=1) and when a positive MSET type was present (Type =1)).

* 1440 data points averaged over Subjects and Replication were used in this regression.

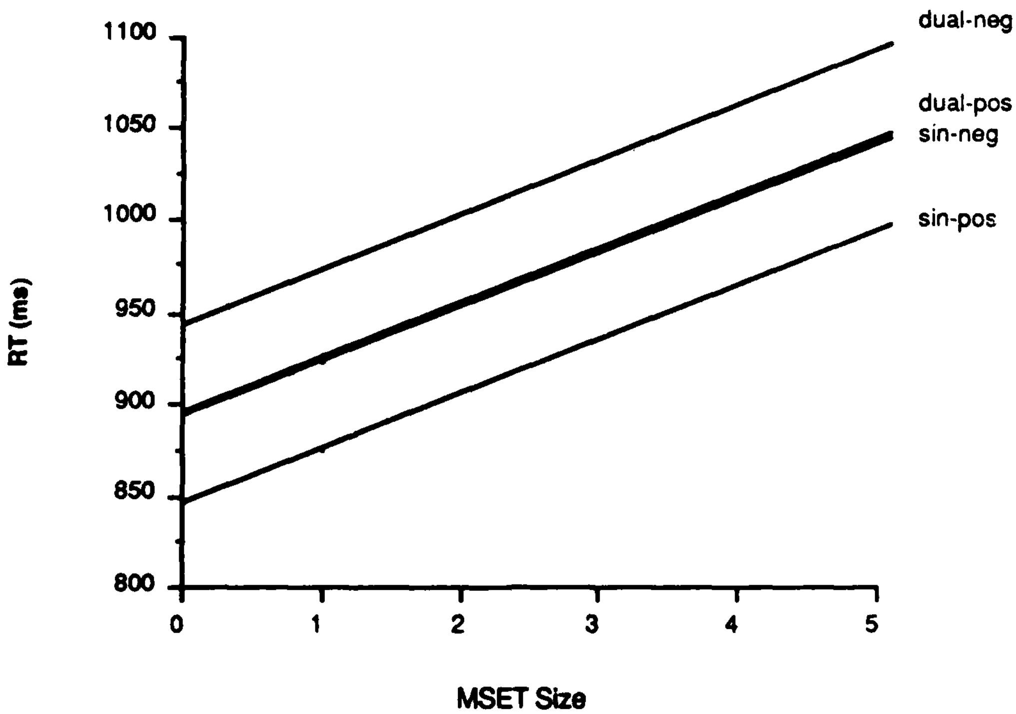


$$RT = 794.71 + 45.21 \text{ MSET} + 23.98 \text{ PT} - 34.00 \text{ Type}$$

$$R^2 = .92^*$$

Figure 24. Least Squares Linear Regression Across MSET for Tracking Level = Medium (categorical variables were included to represent when a dual primary task was present (PT=1) and when a positive MSET type was present (Type=1)).

* 1440 data points averaged over Subjects and Replication were used in this regression.



$$RT = 892.80 + 29.97\ MSET + 50.82\ PT - 47.83\ Type$$

$$R^2 = .66^*$$

Figure 25. Least Squares Linear Regression Across MSET for Tracking Level = High (categorical variables were included to represent when a dual primary task was present (PT=1) and when a positive MSET type was present (Type=1)).

* 1440 data points averaged over Subjects and Replication were used in this regression.

TABLE 14

ANOVA Summary Table for RMS Tracking Error (pixels)

Source	df	SS	MS	F	p
Order	5	9.4513	1.8903	1.36	.3042
Subjects (Order)	12	16.6290	1.3858		
Tracking	2	73.2620	36.6310	326.07	.0001
Tracking * Order	10	0.9338	0.0924	0.82	.6114
Tracking * Subjects (Order)	24	2.6962	0.1123		
Primary Task	1	1.7976	1.7976	43.85	.0001
Primary Task * Order	5	0.1941	0.0388	0.95	.4861
Primary Task * Subjects (Order)	12	0.4920	0.0410		
MSET	3	0.1393	0.0464	3.81	.0181
MSET * Order	15	0.2206	0.0147	1.21	.3119
MSET * Subjects (Order)	36	0.4393	0.0122		
Tracking * Primary Task	2	0.0253	0.0126	1.21	.3144
Tracking * Primary Task * Order	10	0.0728	0.0073	0.70	.7160
Tracking * PT * Subjects (Order)	24	0.2498	0.0104		
Tracking * MSET	6	0.0355	0.0059	0.47	.8292
Tracking * MSET * Order	30	0.4319	0.0144	1.14	.3197
Tracking * MSET * Subj (Order)	72	0.9096	0.0126		
Primary Task * MSET	3	0.0242	0.0081	1.14	.3445
Primary Task * MSET * Order	15	0.0874	0.0058	0.83	.6427
PT * MSET * Subjects (Order)	36	0.2535	0.0070		
Tracking * Primary task * MSET	6	0.0547	0.0091	0.83	.5508
Tracking * PT * MSET * Order	30	0.3240	0.0108	0.98	.5055
Trk * PT * MSET * Subj (Order)	72	0.7912	0.0110		

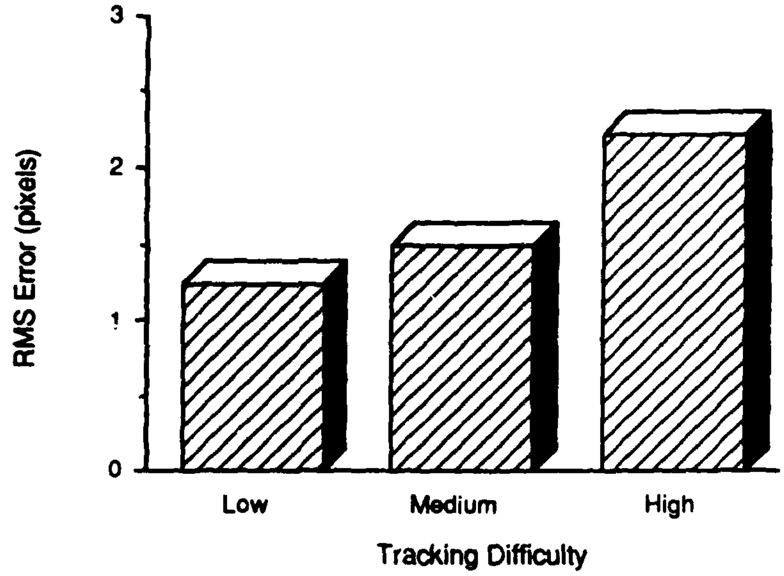


Figure 26. RMS Tracking Error as a Function of Tracking Difficulty.

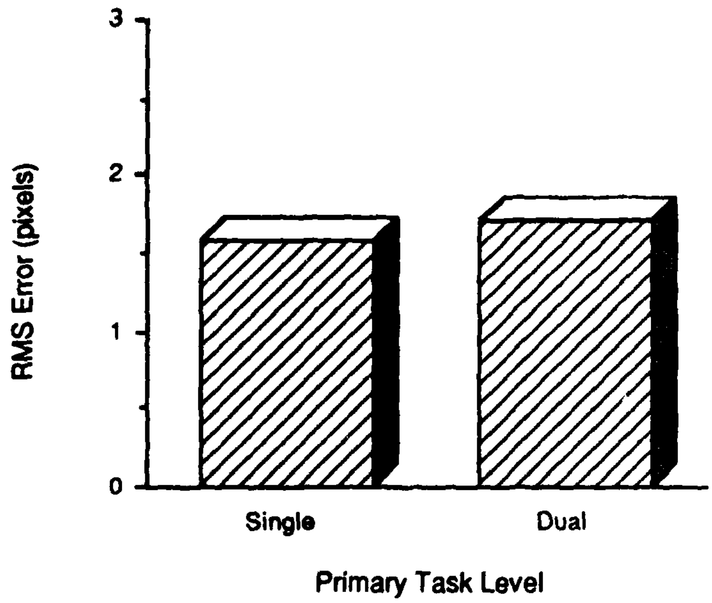


Figure 27. RMS Tracking Error as a Function of Primary Task Level.

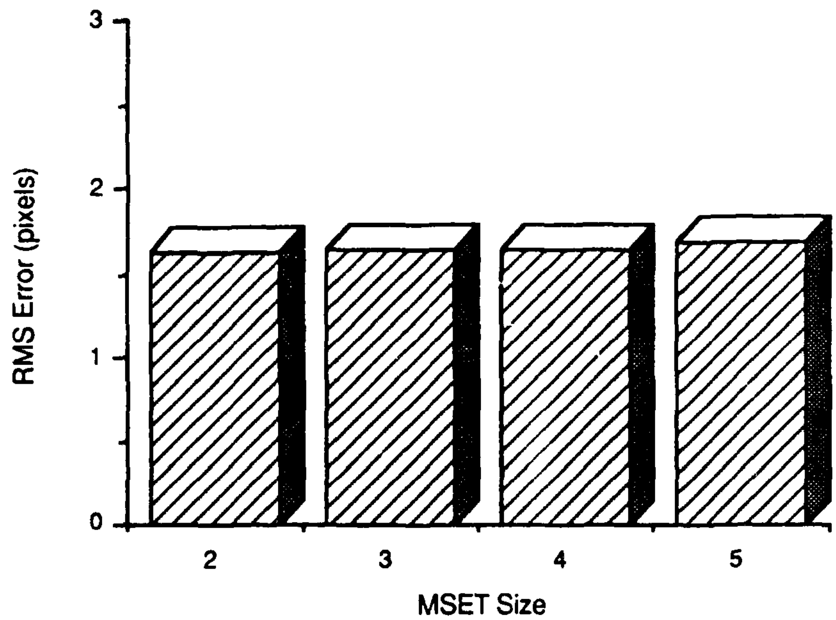


Figure 28. RMS Tracking Error as a Function of MSET Size.

(Table 15) showed all three levels of Tracking difficulty to be significantly different, RMS tracking error increasing as a function of increased Tracking difficulty. A Newman-Keuls Test on MSET size means (Table 16) showed only the two extreme MSET sizes (2 and 5) to have significantly different means, with the MSET size 2 mean being the lowest and the MSET size 5 mean being the highest.

A 3 x 4 x 6 x 8 ANOVA (Tracking difficulty by MSET size by Dual primary task probe position by Order) was conducted on dual primary task RT (Table 17). Two factors were significant as main effects (Order, $p = .0154$ and Probe position, $p < .0001$) and as factors in significant interactions with MSET size.

The MSET size by Probe position interaction ($p = .0009$) is shown in Figure 29. Post hoc Simple-Effect F-Tests (Table 18) revealed that Probe position means were significantly different at all levels of MSET size except for the MSET size of 2. A series of Newman-Keuls Tests was conducted on the means at each significant MSET size. At the MSET size = 3 level (Table 19), no single mean or pair of means was significantly different from all other means. That is, the significant differences at this MSET size level were distributed among all levels of dual primary task probe position. At the MSET size = 4 level (Table 20), the significant differences were located at the extreme values only. Finally, mean RTs for probe position 7 was significantly different from all other means at the MSET size = 5 level (Table 21). The second significant interaction was MSET size by Order ($p = .0323$). This interaction is displayed in Figure 30. Post hoc Simple-Effect F-Tests (Table 22) revealed that Order means were significantly different for all levels of MSET size. Newman-Keuls Tests were conducted on the means at each MSET size (Tables 23, 24, 25, and 26) which showed the differences to be accounted for mainly in the means for Order groups 3 and 2. There appears to be no logical explanation for this result, nor does the result impact other

TABLE 15

Results of Newman-Keuls Test on Tracking Difficulty (Mean RMS Tracking Error (pixels))*

Tracking Difficulty:	Low	Medium	High
Mean Value:	1.242	1.485	2.211

*means with a common line do not differ significantly at $p < .05$.

TABLE 16

Results of Newman-Keuls Test on MSET size (Mean RMS Tracking Error (pixels))*

MSET size:	2	3	4	5
Mean Value:	1.624	1.635	1.652	1.672

*means with a common line do not differ significantly at $p < .05$.

TABLE 17

ANOVA Summary Table for Dual Primary Task Reaction Time

Source	df	SS	MS	F	p
Order	5	31674.5508	6334.9102	4.49	.0154
Subjects (Order)	12	16923.5861	1410.2988		
Tracking	2	402.5754	201.2877	0.70	.5073
Tracking * Order	10	3251.7299	325.1730	1.13	.3828
Tracking * Subjects (Order)	24	6918.5414	288.2726		
MSET	3	657.7230	219.2410	2.59	.0676
MSET * Order	15	2693.0366	179.5358	2.12	.0323
MSET * Subjects (Order)	36	3042.9183	84.5255		
Probe	7	8366.1089	1195.1584	6.93	.0001
Probe * Order	35	6478.3868	185.0968	1.07	.3872
Probe * Subjects (Order)	84	14491.6539	172.5197		
Tracking * MSET	6	610.2905	101.7151	1.29	.2736
Tracking * MSET * Order	30	1403.0153	46.7672	0.59	.9438
Tracking * MSET * Subj (Order)	72	5684.7975	78.9555		
Tracking * Probe	14	1593.6772	113.8341	1.32	.2013
Tracking * Probe * Order	70	6337.2242	90.5318	1.05	.3967
Tracking * Probe * Subj (Order)	168	14510.0053	86.3691		
MSET * Probe	21	4008.5883	190.8852	2.38	.0009
MSET * Probe * Order	105	6140.9244	58.4850	0.73	.9679
MSET * Probe * Subj (Order)	252	20201.9394	80.1664		
Tracking * MSET * Probe	42	2896.8288	68.9721	1.03	.4230
Tracking * MSET * Probe * Order	210	13209.8499	62.9040	0.94	.6986
Trk * MSET * Probe * Subj (Order)	504	33752.5581	66.9694		

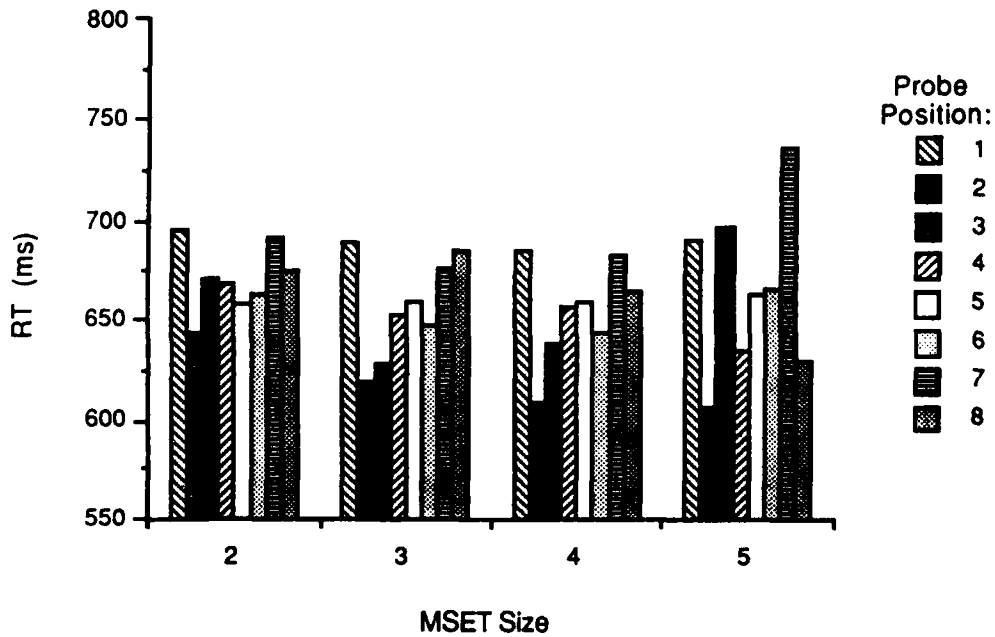


Figure 29. Dual Primary Reaction Time as a Function of MSET Size and Probe Position.

TABLE 18

Results of Simple-Effect F-Tests on Dual Primary Task Probe for Each MSET Size (Reaction Time (ms))

MSET Size	MS_{Probe}	F	p
2	152.7584	1.91	> .05
3	354.3108	4.42	< .001
4	327.1781	4.08	< .001
5	933.5665	11.65	< .001

TABLE 19

Results of Newman-Keuls Test on Dual Primary Task Probe for MSET Size of 3 (Mean Reaction Time (ms)) *

Probe:	2	3	6	4	5	7	8	1
Mean Value:	619.67	627.93	647.48	653.22	659.44	676.52	685.89	689.07

*means with a common line do not differ significantly at $p < .05$.

TABLE 20

Results of Newman-Keuls Test on Dual Primary Task Probe for MSET Size of 4 (Mean Reaction Time (ms))*

Probe:	2	3	6	4	5	8	7	1
Mean Value:	609.22	638.70	643.22	656.59	659.44	665.30	681.70	685.00

*means with a common line do not differ significantly at $p < .05$.

TABLE 21

Results of Newman-Keuls Test on Dual Primary Task Probe for MSET Size of 5 (Mean Reaction Time (ms))*

Probe:	2	8	4	5	6	1	3	7
Mean Value:	607.52	629.81	634.78	663.78	666.15	690.48	695.74	736.07

*means with a common line do not differ significantly at $p < .05$.

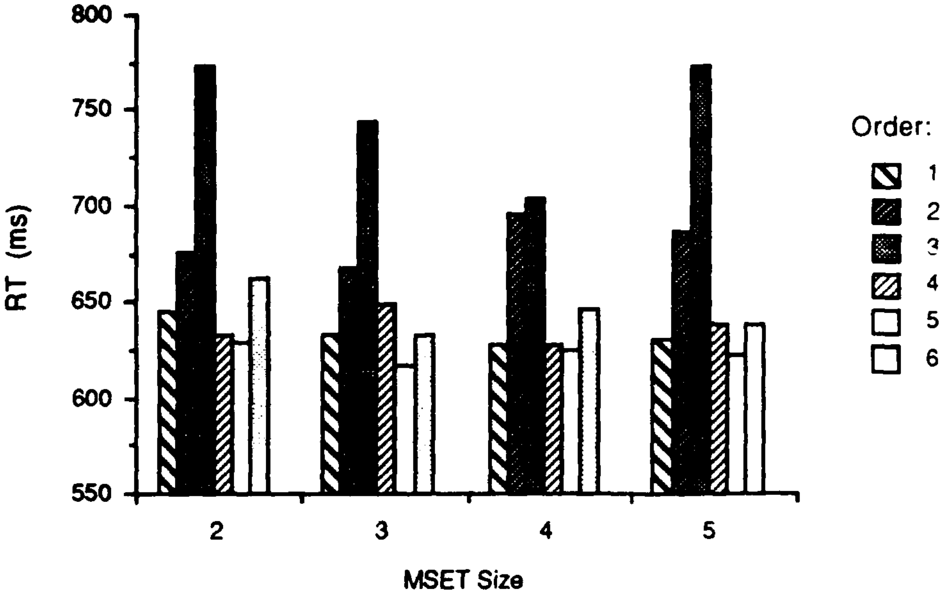


Figure 30. Dual Primary Task Reaction Time as a Function of MSET Size and Order.

TABLE 22

Results of Simple-Effect F-Tests on Order for Each MSET Size (Reaction Time (ms))

MSET Size	MS_{Order}	F	p
2	2054.6468	24.31	< .001
3	1518.2697	17.96	< .001
4	940.0086	11.12	< .001
5	2360.5924	27.93	< .001

TABLE 23

Results of Newman-Keuls test on Order for MSET Size of 2 (Mean Reaction Time (ms)) *

Order:	5	4	1	6	2	3
Mean Value:	630.19	633.58	645.92	663.03	675.36	773.44

*means with a common line do not differ significantly at $p < .05$.

TABLE 24

Results of Newman-Keuls Test on Order for MSET Size of 3 (Mean Reaction Time (ms)) *

Order:	5	6	1	4	2	3
Mean Value:	617.64	632.31	632.47	649.36	668.85	744.14

*means with a common line do not differ significantly at $p < .05$.

TABLE 25

Results of Newman-Keuls Test on Order for MSET Size of 4 (Mean Reaction Time (ms)) *

Order:	5	1	4	6	2	3
Mean Value:	625.56	627.94	628.31	646.89	696.00	704.69

*means with a common line do not differ significantly at $p < .05$.

TABLE 26

Results of Newman-Keuls Test on Order for MSET Size of 5 (Mean Reaction Time (ms)) *

Order:	5	1	4	6	2	3
Mean Value:	623.94	631.53	638.44	639.33	686.56	773.34

*means with a common line do not differ significantly at $p < .05$.

conclusions from the experiment.

No statistically significant effects were found in a 2 x 3 x 4 x 6 ANOVA (MSET type by Tracking difficulty by MSET size by Order) that was performed on dual primary task error percentage. Table 27 is the summary table for this ANOVA. The overall error percentage for dual primary task performance was 1.8 %.

DISCUSSION

This experiment was an initial investigation of the role of short-term memory in operator workload. Specifically, the study was conducted to test the feasibility of reducing short-term memory resource demand as a strategy for the reduction of mental workload in a complex information processing environment. This discussion of results contains four parts: subjective ratings of mental workload, secondary task performance (spare mental capacity) and the Sternberg scanning paradigm, primary task performance, and the role of short-term memory in operator workload. A summary of the experiment and recommendations for future research can be found in Section 4.

Subjective Ratings

The dissociation of subjective ratings from performance ratings has been a topic of increasing concern in the mental workload literature (e.g., Yeh and Wickens, 1988). Because of the prevalence of such dissociation, multiple workload measures, both subjective and objective, were included in this experiment, to obtain multiple ratings of mental workload and to test inter-instrument reliability. An examination of the data in Table 7 leads one to the conclusion that mean SWAT ratings and median MCH values were highly consistent with one

TABLE 27

ANOVA Summary Table for Dual Primary Task Error Percentage

Source	df	SS	MS	F	p
Order	5	158.5842	31.7168	0.80	.5703
Subjects (Order)	12	475.5178	39.6265		
Tracking	2	23.5122	11.7561	0.93	.4085
Tracking * Order	10	264.7337	26.4733	2.09	.0672
Tracking * Subjects (Order)	24	303.5372	12.6474		
MSET	3	27.5364	9.1788	1.02	.3937
MSET * Order	15	176.3384	11.7559	1.31	.2461
MSET * Subjects (Order)	36	322.8677	8.9685		
Type	1	5.4138	5.4138	0.24	.6333
Type * Order	5	107.2954	21.4591	0.95	.4845
Type * Subjects (Order)	12	271.1569	22.5964		
Tracking * MSET	6	46.3059	7.7176	0.98	.4430
Tracking * MSET * Order	30	125.8976	4.1966	0.53	.9706
Tracking * MSET * Subj (Order)	72	565.1361	7.8491		
Tracking * Type	2	7.5515	3.7757	0.31	.7331
Tracking * Type * Order	10	152.7775	15.2778	1.27	.2993
Tracking * Type * Subj (Order)	24	288.1181	12.0049		
MSET * Type	3	34.6137	11.5379	1.15	.3432
MSET * Type * Order	15	168.8386	11.2559	1.12	.3750
MSET * Type * Subj (Order)	36	362.0241	10.0562		
Tracking * MSET * Type	6	83.7156	13.9526	1.39	.2313
Tracking * MSET * Type * Order	30	195.3288	6.5110	0.65	.9065
Trk * MSET * Type * Subj (Order)	72	723.8997	10.0542		

another and with primary task level, but did show some dissociation with other task characteristics and secondary task RT. However, these relationships are better illustrated in Table 8, where it is apparent that MSET size, tracking difficulty, and primary task level are all good predictors of both subjective rating measures when considered as a whole. Again, some dissociation is evident between subjective ratings and secondary task RT. This dissociation occurs primarily in the medium tracking difficulty condition and seems mostly reflective of the fact that the difference between the low and medium tracking difficulty levels was not as great (both subjectively and in terms of RMS tracking error) as the difference between the medium and high levels. It is possible that more objectively different tracking difficulty levels would have yielded less dissociation by increasing the differential effect of tracking difficulty on secondary task RT at the low and medium tracking difficulty levels.

Of the two subjective measures used in this experiment, SWAT proved to be the most sensitive. This is evidenced by the larger range of SWAT scale values used by subjects and the higher correlation of SWAT ratings with secondary task RTs. It is unclear why subjects used the top branch of the MCH decision tree almost to the exclusion of the other branches. One possible interpretation is that the written descriptions used in the MCH scale describe an inappropriately large range of workload situations relative to this experiment. Alternatively, it may be possible that previous results using the MCH scale have been achieved by subjects' use of the final integer values usually included in this scale, rather than the written descriptors and the decision tree logic of the MCH scale. By requiring subjects to step through each scale node one computer screen at a time, and by removing the integer values from the screens, subjects were discouraged in this experiment from "jumping" to a final integer rating and bypassing the logic of the scale.

Another interpretation of this result is that the SWAT rating procedure intruded on MCH ratings (a confounding by instrumentation). That is, since all subjects made three 3-point SWAT ratings per trial in addition to their MCH rating, it is possible that the use of the three 3-point SWAT scales influenced subjects to use only the top 3-point branch of the MCH scale.

If these interpretations are correct, they suggest the need for a careful analysis of the effects various administration techniques may have on the validity of these scales. In particular, if one or both of the scales had intrusive effects on subjective ratings then this finding has implications for the use of multiple subjective rating scales in a within-subjects fashion. Such unanticipated interactions could jeopardize both the sensitivity and validity of the scales used.

Secondary Task Performance

Secondary task performance was hypothesized to be a reflection of the spare mental capacity available to the operator in responding to various levels of short-term memory loading while performing either single or dual primary tasks. The finding that the dual primary task condition and larger MSET size conditions elicited higher secondary task RTs is consistent with the experimental hypotheses. The main effect of Order, however, was unexpected and is not readily explainable. One possible interpretation is that a true Order effect exists in the presentation of tracking difficulty conditions across days. However, a logical pattern does not exist in the data and the Order group 3 is the only group with a statistically different mean. An equally plausible alternative explanation is that the three subjects assigned to group 3 exhibited extreme secondary task RTs by chance and that the three subjects per group was not a large enough number to obtain a stable measure of

performance within each group.

The main effects of MSET type and Tracking difficulty were also anticipated in the design of this experiment and are generally consistent with results reported in the literature. That is, negative MSET responses are typically found to have greater latencies than positive MSET responses and higher levels of tracking difficulty should lead to higher levels of mental workload. However, the interaction of these two effects (Figure 21) was also significant in this experiment. Although this interaction was statistically significant in the ANOVA (p = .039), the post-hoc analysis suggests that this effect was not robust, since MSET type means were significantly different only at the highest level of tracking difficulty.

A least-squares linear regression procedure yielded functions comparable to those reported by Sternberg and others. Specifically, the functions shown in Figures 23, 24, and 25 are parallel lines, increasing as linear functions of MSET size. Sternberg's serial, exhaustive scanning model predicts an increase in RT as a linear function of MSET size due to a longer amount of central processing time needed to complete the mental recognition scan. The slopes of the lines in these figures (29.77 ms to 45.21 ms per MSET digit) are comparable to those reported elsewhere (e.g., Smith and Langolf, 1981) and the fit of the lines is good (R^2 = .64 to .92). One irregularity does exist in these slopes. In the Sternberg scanning model, the slope of the obtained function is generally interpreted as a reflection of the efficiency of the mental scan, or the inverse of the capacity of working memory (Cavanaugh, 1972). A low slope indicates an efficient mental scan. Therefore, one would expect the highest slope in this experiment to be found for the highest level of tracking difficulty, assuming this task required central processing resources. Such was not the case, as the lowest slope was found for the highest level of tracking difficulty. Rather, a

consideration of the y-intercepts of these functions indicates that, according to the Sternberg model, tracking difficulty added mainly to input/output demand. The higher y-intercepts found for the negative MSET condition at tracking levels of medium and high difficulty are also accounted for by Sternberg's model as the result of elevated response/output delay and are consistent with the finding of other researchers (e.g., Wickens et al., 1986). Likewise, multiple resource theory would predict elevated RTs in the dual versus single primary task conditions due to increased perceptual demand in the dual primary task condition.

Secondary task error percentage was not as sensitive a performance measure as secondary task RT. While the ANOVA on this measure did yield one significant main effect (Primary task level) the primary utility of this measure seems to be one of validating the appropriateness of the RT data for the Sternberg scanning model, which assumes correct responses (accurate scans). Since the average percentage of errors committed (5.1 %) was comparable or lower than data reported in other analyses using the Sternberg model (e.g., Wickens, Moody, and Vidulich, 1985), the data seem appropriately applied in this regard.

Primary Task Performance

In the secondary task paradigm, subjects are instructed to direct their attentional resources to the primary task with priority over secondary task performance. Given the difficulty of voluntary allocation of resources in a complex informational environment and the dual nature of the primary task in this experiment, it is not surprising that the addition of the dual task and increased levels of tracking difficulty and MSET size had detrimental effects on RMS tracking error. The effect of MSET size indicates that secondary task intrusion on primary task performance occurred. This finding adds to the popular

criticism of secondary task paradigms; that is, subjects are unable to allocate their attentional resources in an exact manner. When secondary task intrusion occurs, what is being measured as secondary task performance may not be a true reflection of spare mental capacity.

MSET size also intruded on dual primary task RT in interactions with Order and dual primary task Probe position. Again, Order group 3 accounted for the majority of the Order effect in the first interaction and one must question if this is a true "Order" effect at all. Possibly the most legitimate conclusion to draw from these intrusion effects is that resource allocation strategies used by subjects may be both complex and context dependent.

The analysis of error percentage for dual primary task performance showed this measure to be less sensitive than dual primary task RT. This may be due in part to the relatively low number of errors committed while performing this task (1.8 %).

The Role of Short-Term Memory in Operator Workload

It is clear that increased short-term memory loading (MSET size) produced corresponding increases in both subjective and objective measures (secondary task RT) of mental workload. Furthermore, good linear estimates were drawn to describe secondary task RT as a function of MSET size, MSET type, Primary task level, and Tracking difficulty.

The results of this initial investigation indicate that short-term memory does play a significant role in the composition of the multidimensional construct of mental workload. Furthermore, this role can be described in both qualitative and quantitative terms using validated models of human information processing such as the Sternberg scanning model.

IV. SUMMARY AND CONCLUSIONS

INITIAL EXPERIMENT

Summary

A combined dual and secondary task paradigm was used to evaluate the effects of short-term memory loadings on mental workload. Memory loading was varied by increasing stimulus set size in a visual Sternberg recognition memory paradigm. Subjects performed a continuous compensatory tracking task and a visual choice response task in the dual primary task conditions. Secondary task performance was Sternberg short-term recognition memory reaction time. Subjective ratings of mental workload were made using computerized versions of the Modified Cooper-Harper scale and SWAT.

Evaluation

Analyses indicate that secondary task performance degraded with elevated short-term memory loading, primary task level, and as an interaction of increased tracking difficulty and probe membership in MSET. In this regard, the Sternberg scanning task was an effective instrument for the manipulation of mental workload levels. A linear regression performed on secondary task RTs yielded linear functions by MSET size which are consistent with those reported in the literature; that is, RTs increased as MSET size increased in a linear fashion with slopes between 29 ms and 45 ms per MSET digit. In addittion, y-intercepts were higher for linear functions in the dual versus single primary task condition and the negative versus positive MSET condition, as expected.

Increased MSET (short-term memory) loading produced corresponding increases in subjective ratings of mental workload. This finding is consistent with the hypothesis of

Yeh and Wickens (1988) that working memory demands are a major factor of influence in subjective workload ratings. However, while SWAT and MCH ratings were highly correlated, only SWAT ratings were significantly correlated with secondary task RT. This seems to be primarily due to subjects' use of a limited range of the MCH scale.

An intrusion analysis on dual primary task performance revealed two second-order interactions and a main effect involving MSET size, indicating a degree of secondary task intrusion on both tracking performance and visual choice RT.

RECOMMENDATIONS FOR FUTURE RESEARCH

Assessment tools

A consideration of the performance of the MCH and SWAT rating scales in this experiment leads to the conclusion that SWAT is the more sensitive technique in this research paradigm. Perhaps more important is the possibility that the two scales themselves confounded the subjective ratings when used as within-subjects factors. This possibility calls into question the validity of subjective scale ratings when used in this fashion. Further research should explore alternative subjective rating techniques (e.g., free-modulus magnitude estimation) while retaining relatively well proven tools such as the SWAT scaling technique in the design as a between-subjects factor.

The Sternberg scanning paradigm proved to be both an effective assessment strategy for mental workload as well as a viable means by which short-term memory demands may be manipulated in a secondary task paradigm. Furthermore, Sternberg short-term memory loading did not lead to the large dissociations which are often reported between subjective and performance measures of mental workload. However, the discovery of secondary task

intrusion in this study could be considered as qualifying the assumption that subjects are able to precisely allocate mental resources in the secondary task paradigm.

Reduction of mental workload

Given the conclusion that short-term memory loading does play a significant role in operator mental workload, strategies for the reduction of short-term memory demands may also be considered as viable strategies for the reduction of workload demands. Such strategies discussed in Section 1 of this report include:

1. Grouping (Baddely, 1982; Frick, 1984; Miller, 1956);
2. Hierarchical organization (Loftus and Loftus, 1976);
3. Elimination of distractor tasks (Loftus and Loftus, 1976);
4. Release from Proactive Interference (Loftus and Loftus, 1976);
5. Rehearsal;
6. Redundant visual cuing (Simon, 1984);
7. Adjunctive rehearsal mechanisms (Reisberg et al., 1984);
8. Dual modality storage (Frick, 1984, 1985);
9. Optimization of stimulus class (Cavanaugh, 1972; Loftus et al., 1979);
10. Automated information management (Goodstein, 1981; Kuperman and Wilson, 1986; Frick et al., 1986; Rassmussen, 1981, 1986, 1987).

Specific recomendations for future research to be discussed will target strategies 9 and 10. The domain of application for this recommended research will be the crewstation of the advanced conceptual bomber within the context of the RT mission.

Application to manned bomber systems

The role of the manned bomber in the execution of the RT mission was described by Frick, Hoover, Campbell, Cotton, Aaranson, Kuperman, and Wilson (1986). The typical RT mission scenario presents an extreme case of elevated operator mental workload due to the long mission duration; large number of possible targets; uncertainty of target presence, activity, or location; and the necessity for low altitude flight through a rapidly changing environment (Frick et al., 1986). The manned bomber is well suited for use in such scenarios because of the inherent flexibility afforded by the inclusion of a reasoning crew held in-the-loop. Unfortunately, operators in-the-loop also contribute their human frailties to the process (i.e., sensory and information processing limitations). In particular, operator short-term memory limitations may be partially responsible for elevated aircrew mental workload.

The results from this initial experiment support the viability of applying the short-term memory concept to the information environment of the advanced conceptual bomber to effect the reduction of mental workload. Future research should investigate specific implementation strategies and parameters.

Kuperman and Wilson (1986) suggested that a principal means by which mental workload may be reduced and situational awareness increased in the advanced bomber crewstation is the introduction of expert system technology. Ben-Bassat and Freedy (1982) used the term "decision support systems" to describe the role of computer-based expert systems in reducing decision uncertainty by increasing situational awareness. The implementation of an expert decision support system to the air crewstation might be most effectively accomplished by (1) structuring the system to minimize short-term memory loading (Rassmussen, 1981), (2) using system aiding to prevent short-term memory recall errors in rule-based behavior

(Rassmussen, 1987), and (3) avoiding unnecessarily rigid information presentation (Goodstein, 1981).

Research outline

A series of experiments is suggested to apply the results of this literature review and initial experiment. The following research objectives should be considered:

1. Test the workload reduction utility of short-term memory aiding in an expert system environment;
2. Test the workload reduction utility of short-term memory aiding for Rassmussen's (1983) rule-based class of behavior;
3. Describe the effect of short-term memory aiding on operator situational awareness;
4. Investigate various augmentation strategies for information presentation within an automated information management system; and
5. Demonstrate how the short-term memory concept might be applied in a crew station design.

Several experiments would be necessary to accomplish the stated objectives; one experiment to accomplish objective 3, one experiment to accomplish objective 4, and one experiment to follow-up on results from the first two experiments. Objectives 1, 2 and 5 would be integrated into all three experiments.

The experimental venue would be abstracted from the RT mission scenario and might involve operator's rule-based processing of battle management information pertinent to the system defense mode. In these three experiments subjects could be tasked with the mental (short-term memory) inventory (identification and location) and response to various threats (e.g., Air

Interceptors (AI), Surface-to-Air Missiles (SAM), or Airborne Warning and Control Systems (AWACS)). Experiment 1 might investigate environmental situation awareness information (e.g., heading, wind speed, visibility, terrain) in the context of aircrew communication or Integration of sensor data with database information .

Experiment 2 could be designed to test augmentation of threat symbology in accordance with objective 4 in this research. One strategy to accomplish this might include optimization of stimulus class. Loftus et al. (1979) discussed the short-term memory of numerical information in ground controller/pilot communication and the variation of stimulus class for the encoding of numerical information. Cavanagh (1972) reported higher recall memory spans and shorter Sternberg scanning times (i.e., larger memory capacities) for some stimulus classes (e.g., digits) over others (e.g., letters). Memory loading for digit, color, letter, shape, and word stimuli and their contribution to mental workload could be evaluated. Based on these results, crew station display formatting recommendations would be generated.

Experiment 3 could be reserved for integrating or replicating pertinent effects from the first two experiments.

These experiments could be conducted in an abstracted version of Kuperman and Wilson's (1986) SABER crew station. A single color display surface representing, for example, the "current action" or horizontal situation information display could be driven by a personal computer with joystick cursor control and multipurpose keyboard input. This apparatus would represent a partial, abstracted version of the SABER crew station.

Together, these three experiments represent an opportunity to apply the STM concept in a direct manner to the anticipated

problem of elevated operator mental workload in the advanced bomber.

REFERENCES

Allen, C.K., 1983, Short-term memory for colors and color names. Psychological Reports, 53, 579-582.

Allen, C.K., 1984, Short-term memory for colors and color names in the absence of vocalization. Perceptual and Motor Skills, 59, 263-266.

Atkinson, R.C. and Shiffrin, R.M., 1968, Human memory: A proposed system and its control processes. In K.W. Spence and J.J. Spence (Eds.), The psychology of learning and motivation: Advances in research and theory (Vol. 2) (pp. 89-195). New York: Academic Press.

Atkinson, R.C. and Shiffrin, R.M., 1971, The control of short-term memory. Scientific American, 225, 82-90.

Armstrong Aerospace Medical Research Laboratory (Draft), July 1987, Subjective Workload Assessment Technique (SWAT): A users guide. Harry G. Armstrong Aerospace Medical Research Laboratory. Dayton, OH: Wright-Patterson AFB.

Baddeley, A.D., 1966, Short-term memory for word sequences as a function of acoustic, semantic, and formal similarity. Quarterly Journal of Experimental Psychology, 18, 362-365.

Baddeley, A.D., 1972, Selective attention and performance in dangerous environments. British Journal of Psychology, 63 (4), 537-546.

Baddeley, A.D., 1976, The psychology of memory. New York: Basic Books, Inc.

Baddeley, A.D., 1982, Your memory: A user's guide. New York: Macmillan Publishing Co.

Baddeley, A.D., 1984, The fractionation of human memory. Psychological Medicine, 14, 259-264.

Baddeley, A.D., Grant, S., Wight, E., and Thomson, N., 1975, Imagery and visual working memory. In P. Rabbitt and S. Dornic (Eds.), Attention and performance V (pp. 205-217). New York: Academic Press.

Baddeley, A.D. and Hitch, G., 1974, Working memory. In G.H. Bower (Ed.), The psychology of learning and motivation:

Advances in research and theory (Vol. 8, pp. 47-89). New York: Academic Press.

Baddeley, A.D. and Lieberman, K., 1980, Spatial working memory. In R.S. Nickerson (Ed.), Attention and performance VIII (pp. 521-539). Hillsdale NJ: Erlbaum.

Baddeley, A.D., Thomson, N., and Buchanan, M., 1975, Word length and the structure of short-term memory. Journal of Verbal Learning and Verbal Behavior, 14, 575-589.

Ben-Bassat, M. and Freedy, A., 1982, Knowledge requirements and management in expert decision support systems for (military) situation assessment. IEEE Transactions on Systems, Man, and Cybernetics, SMC-12 (4), 479-490.

Berliner, C., Angell, D., and Shearer, J.W., 1964, Behaviors, measures, and instruments for performance evaluation in simulated environments. In Proceedings of the Symposium and Workshop on the Quantification of Human Performance. Albuquerque, NM.

Boyd, S. P., 1983, Assessing the validity of SWAT as a workload measurement instrument. In Proceedings of the Human Factors Society 27th Annual Meeting (pp. 124-128). Santa Monica, CA: Human Factors Society.

Brown, J., 1958, Some tests of the decay theory of immediate memory. Quarterly Journal of Experimental Psychology, 10, 12-21.

Bundeson, C., Pedersen, L.F., and Larsen, A., 1984, Measuring efficiency of selection from briefly exposed visual displays: A model for partial report. Journal of Experimental Psychology: Human Perception and Performance, 10 (3), 329-339.

Casali, J. G., 1982, A sensitivity/intrusion comparison of mental workload estimation techniques using a simulated flight task emphasizing perceptual piloting behaviors. Unpublished disertation, VPI & SU, Blacksburg, VA 24061.

Casali, J. G. and Wierwille, W. W., 1983, A comparison of rating scale, secondary-task, physiological, and primary task load estimation techniques in a simulated flight task

emphasising communications load. Human Factors, 25 (6), 623-641.

Casali, J. G. and Wierwille, W. W., 1984, On the measurement of pilot perceptual workload: a comparison of assessment techniques addressing sensitivity and intrusion issues. Ergonomics, 27 (10), 1033-1050.

Case, R., Kurland, D.M., and Goldberg, J., 1982, Operational efficiency and the growth of short-term memory span. Journal of Experimental Child Psychology, 33, 386-404.

Cavanagh, J.P., 1972, Relation between the immediate memory span and the memory search rate. Psychological Review, 79, 525-530.

Cermak, L.S., 1982, The long and short of it in amnesia. In L.Cermak (Ed.), Human memory and amnesia (pp. 43-59). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Chase, W.G. and Calfee, R.C., 1969, Modality and similarity effects in short-term recognition memory. Journal of Experimental Psychology, 81, 510-514.

Cooper, G. E. and Harper, R. P., 1969, The use of pilot rating in the evaluation of aircraft handling qualities (Tech Report AGARD 567). Ames Research Center, Moffett Field: CA.

Crabtree, M. S., Bateman, R. P. and Acton, W. H., 1984, Benefits of using objective and subjective workload measures. In Proceedings of the Human Factors Society 28th Annual Meeting (pp. 950-953). Santa Monica, CA: Human Factors Society.

Craik, F.I., 1968, Two components in free recall. Journal of Verbal Learning and Verbal Behavior, 7, 996-1004.

Craik, F.I., 1970, The fate of primary memory items in free recall. Journal of Verbal Learning and Verbal Behavior, 9, 143-148.

Craik, F.I. and Lockhart, R.S., 1972, Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 11, 671-684.

Craik, F.I. and Watkins, M.J., 1973, The role of rehearsal in short-term memory. Journal of Verbal Learning and Verbal

Behavior, 12, 599-607.

Crannell, C.W. and Parrish, J.M., 1957, A comparison of immediate memory span for digits, letters, and words. The Journal of Psychology, 44, 319-327.

Crowder, R.G. and Morton, J., 1969, Precategorical acoustic storage (PAS). Perception and Psychophysics, 5, 365-373.

Curry, R.E., 1981, A model of human fault detection for complex dynamic processes. In J. Rasmussen and W.B. Rouse (Eds.), Human detection and diagnosis of system failures (pp. 171-183), New York: Plenum Press.

Daneman, M. and Carpenter, P.A., 1980, Individual differences in working memory and reading. Journal of Verbal Learning and Verbal Behavior, 19, 450-466.

Dempster, F. N., 1981, Memory span: Sources of individual and developmental differences. Psychological Bulletin, 89 (1), 63-100.

Eggemeier, F. T., 1984, Workload metrics for system evaluation. In Proceedings of the NATO Defense Research Group Panel VIII Workshop on Applications of Systems Ergonomics to Weapon System Development (Vol. 1). Tech Report, pp. C5-C20. Royal Military Colleage of Science, Shrivenham, UK.

Eggemeier, F. T., Crabtree, M. S. and LaPointe, P. A., 1983, The effect of delay report on subjective workload ratings. In Proceedings of the Human Factors Society 27th Annual Meeting (pp. 139-143). Santa Monica, CA: Human Factors Society.

Eggemeier, F. T., Crabtree, M.S., Zingg, J. J., Reid, G. B. and Shingledecker, C. A., 1982, Subjective workload assessment in a memory update task. In Proceedings of the Human Factors Society 26th Annual Meeting (pp. 643-647). Santa Monica, CA: Human Factors Society.

Eggemeier, F. T. and Stadler, M. A., 1984, Subjective workload assessment in a spatial memory task. In Proceedings of the Human Factors Society 28th Annual Meeting (pp. 680-684). Santa Monica, CA: Human Factors Society.

Eggleston, R. G. and Quinn, T. J., 1984, A preliminary evaluation of a projective workload assessment procedure. In Proceedings of the Human Factors Society 28th Annual Meeting (pp. 695-699). Santa Monica, CA: Human Factors Society.

Eisen, P.S. and Hendy, K.C., 1987, A preliminary examination of mental workload, its measurement and prediction. Downsview, Ontario: Defense and Civil Institute of Environmental Medicine.

Elliot, D. and Jones, R., 1984, Short-term memory for duration of movement. Perceptual and Motor Skills, 58, 331-334.

Ellis, N.C. and Hennelly, R.A., 1980, A bilingual word-length effect: Implications for intelligence testing and the relative ease of mental calculation in Welsh and English. British Journal of Psychology, 71, 43-51.

Eprath, A.R. and Young, L.R., 1981, Monitoring vs. man-in-the-loop detection of aircraft control failures. In J. Rasmussen and W.B. Rouse (Eds.), Human detection and diagnosis of system failures (pp. 143-154), New York: Plenum Press.

Erdelyi, M.H. and Becker, J., 1974, Hypermnesia for pictures: Incremental memory for pictures but not words in multiple recall trials. Cognitive Psychology, 6, 159-171.

Fitzgerald, P. and Broadbent, D.E., 1985, Order of report and the structure of temporary memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11 (2), 217-228.

Firth, P.A., 1973, Psychological factors influencing the relationship between cardiac arrhythmia and mental load. Ergonomics, 16, 5-16.

Frick, R.K., Hoover, R., Campbell, B., Cotton, F., Aaranson, J., Kuperman, G.G., Wilson, D., 1986, RT mission scenario description. Harry G. Armstrong Aerospace Medical Research Laboratory, Wright Patterson AFB, OH. (SECRET).

Frick, R.W., 1984, Using both an auditory and a visual short-term store to increase digit span. Memory and Cognition,

12 (5), 507-514.

Gaddes, W.B. and Brady, L.R., 1981, A method for optimizing human performance in detecting and diagnosing mission avionics faults. In J. Rasmussen and W.B. Rouse (Eds.), Human detection and diagnosis of system failures (pp. 523-538), New York: Plenum Press.

Gilmartin, K.J., Newell, A., and Simon, H.A., 1976, A program modeling short-term memory under strategy control. In C.N. Cofer (Ed.), The structure of human memory (pp. 15-30). San Francisco: W.H. Freeman.

Goodstein, L.P., 1981, Discriminative display support for process operators. In J. Rasmussen and W.B. Rouse (Eds.), Human detection and diagnosis of system failures (pp. 433-449), New York: Plenum Press.

Gopher, D., 1982, A selective attention test as a predictor of success in flight training. Human Factors, 24 (2), 173-183.

Gregg, V.H., 1986, Introduction to human memory. London: Routledge and Kegan Paul Ltd.

Hancock, P. A., Meshkati, N. and Robertson, M. M., 1985, Physiological reflections of mental workload. Aviation, Space, and Environmental Medicine, 56, 1110-1114.

Hitch, G., 1978, The role of short-term working memory in mental arithmetic. Cognitive Psychology, 10, 302-323.

Hitch, G., 1983, Short-term memory processes in humans and animals. In A. Mayes (Ed.), Memory in animals and humans (pp. 177-202), Cambridge, Great Britain: Van Nostrand Reinhold (UK) Co. Ltd.

Hitch, G., 1984, Working memory. In Psychological Medicine, 14, 265-271.

Hitch, G., 1985, Short-term memory and information processing in humans and animals: Towards an integrative framework. In L. Nilsson and T. Archer (Eds.), Perspectives on learning and memory (pp. 119-136). Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

Johannsen, G., Pfender, C. and Stein, W., 1976, Human

performance and workload in simulated landing approaches with autopilot failures. In T. B. Sheridan and G. Johannsen (Eds.), Monitoring behavior and supervisory control. New York: Plenum Press.

Kantowitz, B.H. and Sorkin, R.D., 1983, Human factors: Understanding people-system relationships. New York: John Wiley & Sons.

Keppel, G. and Underwood, B.J., 1962, Proactive inhibition in short-term retention of single items. Journal of Verbal Learning and Verbal Behavior, 1, 153-161.

Klapp, S.T. and Philipoff, A., 1983, Short term memory limits in performance. Turning the tide of technology: Proceedings of the Human Factors Society, 27th annual meeting (pp. 452-454). Santa Monica, CA: Human Factors Society.

Kuperman, G.G. and Wilson, D.L., 1986, A human-centered technology for advanced bomber crew stations. (Tech. Report No. AAMRL-TR-86-052). Harry G. Armstrong Aerospace Medical Research Laboratory, Wright Patterson AFB, OH.

Lavach, J.F., 1973, The effect of arousal on short and long-term memory. The Journal of Educational Research, 67 (3), 131-133.

Loftus, G.R., Dark, V.J., and Williams, D., 1979, Short-term memory factors in ground controller/pilot communication. Human Factors, 21 (2), 169-181.

Loftus, G.R. and Loftus, E.F., 1976, Human memory: The processing of information. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc.

McCloy, T., Derrick, W. and Wickens, C., 1983, Workload assessment metrics: What happens when they dissociate? In Second Aerospace Behavioral Engineering Technology Conference Proceedings (p. 132). Warrendale, PA.

McKendry, J.M. and Hurst, P.M., 1971, Adaptation to speed stress in an immediate memory task. Human Factors, 13 (6), 543-552.

Melton, A.W., 1963, Implications of short-term memory for a

general theory of memory. Journal of Verbal Learning and Verbal Behavior, 2, 1-21.

Mewaldt, S.P., Hinrichs, J.V., and Ghoneim, M.M., 1983, Diazepam and memory: Support for a duplex model of memory. Memory and Cognition, 11 (6), 557-564.

Miller, G.A., 1956, The magical number seven, plus or minus two: Some limits on our capacity for processing information. The Psychological Review, 63 (2), 81-97.

Moray, N., 1967, Where is capacity limited? A survey and a model. In A. Sanders (Ed.), Attention and performance I (pp. 84-92). Amsterdam: North-Holland.

Morton, J., 1969, Interaction of information in word recognition. Psychological Review, 76, 165-178.

Murdock, B.B., 1963, Short-term memory and paired-associate learning. Journal of Verbal Learning and Verbal Behavior, 2, 320-328.

Nilsson, L., Ohlsson, K., and Ronnberg, J., 1977, Capacity differences in processing and storage of auditory and visual input. In S. Dornick (Ed.), Attention and Performance (Vol. VI). Hillsdale, NJ: Lawrence Erlbaum.

Notestine, J. C., 1984, Subjective workload assessment and effect of delayed ratings in a probability monitoring task. In Proceedings of the Human Factors Society 28th Annual Meeting (pp. 685-689). Santa Monica, CA: Human Factors Society.

Nygren, T. E., 1982, Conjoint measurement and conjoint scaling: A users guide (Tech. Report No. 82-22). Wright-Patterson AFB, OH.

Peterson, L.R. and Peterson, M.J., 1959, Short-term retention of individual verbal items. Journal of Experimental Psychology, 58 (3), 193-198.

Pew, R.W., 1979, Secondary tasks and workload measurement. In N. Moray (Ed.), Mental workload, its theory and measurement (pp.23-28). New York: Plenum Press.

Posner, M.I., 1964, Rate of presentation and order of recall in immediate memory. British Journal of Psychology, 55,

303-306.

Posner, M.I. and Keele, S.W., 1967, Decay of information from a single letter. Science, 158, 137-139.

Posner, M.I. and Rossman, E., 1965, Effect of size and location of informational transforms upon short-term retention. Journal of Experimental Psychology, 70, 496-505.

Poulton, E.C., 1974, Tracking skill and manual control. New York, Academic Press.

Proctor, J.H., 1969, Identification of data display requirements in command control. In I.J. Gabelman (Ed.), Displays for command and control centers, NATO AGARD Conference Proceedings No. 23. New York: Rome Air Development Center.

Puff, C.R. (Ed.)., 1982, Handbook of research methods in human memory and cognition. New York: Academic Press.

Rahimi, M. and Wierwille, W. W., 1982, Evaluation of the sensitivity and intrusion of workload estimation techniques in piloting tasks emphasizing mediational behavior. In Proceedings of the 1982 International Conference on Cybernetics and Society (pp. 593-597). New York: I.E.E.E.

Rasmussen, J., 1981, Models of mental strategies in process plant diagnosis. In J. Rasmussen and W.B. Rouse (Eds.), Human detection and diagnosis of system failures (pp. 241-258), New York: Plenum Press.

Rasmussen, J., 1983, Skills, rules, and knowledge; Signals, signs, and symbols, and other distinctions in human performance models. IEEE Transactions on Systems, Man, and Cybernetics, 13 (3), 257-266.

Rasmussen, J., 1986, Information processing and human-machine interaction: An approach to cognitive interaction. New York: North-Holland.

Rasmussen, J., 1987, Cognitive control and human error mechanisms. In J. Rasmussen, K. Duncan, and J. Leplat (Eds.), New technology and human error (pp. 53-61). Chichester, G.B: John Wiley and Sons.

Reid, G. B., 1985, Current status of the development of the

subjective workload assessment technique. In Proceedings of the Human Factors Society 29th Annual Meeting (pp. 220-223). Santa Monica, CA: Human Factors Society.

Reid, G. B., Shingledecker, C. A. and Eggemeier, F. T., 1981, Application of conjoint measurement to workload scale development. In Proceedings of the Human Factors Society 25th Annual Meeting (pp. 522-525). Santa Monica, CA: Human Factors Society.

Reisberg, D. Rappaport, I., and O'Shaughnessy, M., 1984, Limits of working memory: The digit digit-span. Journal of Experimental Psychology: Learning, Memory, and Cognition, 10 (2), 203-221.

Rohmert, W., 1971, An international symposium on objective assessment of work load in air traffic control tasks. Ergonomics, 14 (5), 545-547.

Rolfe, J.M., 1971, The secondary task as a measure of mental load. In W.T. Singleton, J.G. Fox, and D. Whitfield (Eds.), Measurement of man at work (pp. 135-148). London: Taylor and Francis Ltd.

Rolfe, J. M., 1976, The measurement of human response in man-vehicle control situation. In T. B. Sheridan and G. Johannsen (Eds.), Monitoring behavior and supervisory control. New York: Plenum Press.

Salame, P. and Baddeley, A.D., 1982, Disruption of short-term memory by unattended speech: Implications for the structure of working memory. Journal of Verbal Learning and Verbal Behavior, 21, 150-164.

Sanders, M. S. and McCormick, E. J., 1987, Human factors engineering and design (6th ed.). New York: McGraw-Hill.

Santrock, J.W., 1986, Processes of memory and thought. In Psychology: The science of mind and behavior (pp. 255-278). Dubuque, IA: Wm. C. Brown.

Shaffer, W.O. and Shiffrin, R.M., 1972, Rehearsal and storage of visual information. Journal of Experimental Psychology, 92, 292-296.

Shapiro, S.R. and Erdelyi, M.H., 1974, Hypermnesia for

pictures but not words. Journal of Experimental Psychology, 103 (6), 1218-1219.

Sheridan, T.B., 1981, Diagnostic behavior in nuclear power plants. In J. Rasmussen and W.B. Rouse (Eds.), Human detection and diagnosis of system failures (pp. 19-35), New York: Plenum Press.

Sheridan, T. and Simpson, R., 1979, Toward the definition and measurement of the mental workload of transport pilots (Tech. Report No. R 79-4). Massachusetts Institute of Technology, Flight Transportation Laboratory, Cambridge, MA.

Shulman, H.G., 1970, Encoding and retention of semantic and phonemic information in short-term memory. Journal of Verbal Learning and Verbal Behavior, 9, 499-508.

Simon, J.R., 1984, The effect of redundant cues on retrieval time. Human Factors, 26 (3), 315-321.

Skipper, J. H., Rieger, C. A. and Wierwille, W. W., 1986, Evaluation of decision-tree rating scales for mental workload estimation. Ergonomics, 29 (4), 585-599.

Smith, P.J. and Langolf, G.D., 1981, The use of Sternberg's memory scanning paradigm in assessing effects of chemical exposure. Human Factors, 23 (6), 701-708.

Sperling, G., 1960, The information available in brief visual presentations. Psychological Monographs, 74 (Whole No. 11).

Sperling, G., 1963, A model for visual memory tasks. Human Factors, 5, 19-31.

Sperling, G., 1967, Successive approximations to a model for short-term memory. Acta Psychologica, 27, 285-292.

Sternberg, S., 1966, High speed scanning in human memory. Science, 153, 652-654.

Sternberg, S., 1969 a, The discovery of processing stages: extensions of Donders' method. Acta Psychologica, 30, 276-315.

Sternberg, S., 1969 b, Memory-scanning: Mental processes revealed by reaction time experiments. American Scientist,

57, 421-457.

Stone, G., Gulick, R.K., and Gabriel, R.F., 1987, Use of task timeline analysis to assess crew workload. In A.H. Roscoe (Ed.), NATO AGARDograph No. 282 (pp. 15-31). Bedfordshire, UK.

Thompson, D.A., 1981, Commercial aircrew detection of system failures: State of the art and future trends. In J. Rasmussen and W.B. Rouse (Eds.), Human detection and diagnosis of system failures (pp. 37-48), New York: Plenum Press.

Tulving, E., 1968, Theoretical issues in free recall. In T.R. Dixon and D.L. Horton (Eds.), Verbal behavior and general behavior theory. Englewood Cliffs, NJ: Prentice-Hall.

Underwood, B.J., 1983, Attributes of a memory. In B.J. Underwood (Ed.), Attributes of memory (pp 56-71). Glenview, Il: Scott, Foresman and Co.

Ursin, H. and Ursin, R., 1979, Physiological indicators of mental load. In N. Moray (Ed.), Mental workload, its theory and measurement (pp. 340-365). New York: Plenum Press.

Vallar, G. and Baddeley, A.D., 1984, Fractionation of working memory: Neuropsychological evidence for a phonological short-term store. Journal of Verbal Learning and Verbal Behavior, 23, 151-161.

Van der Heijden, A.H., 1981, Short-term visual information forgetting. London: Routledge and Kegan Paul Ltd.

Vidulich, M. A. and Tsang, P. S., 1985, Assessing subjective workload assessment: A comparison of SWAT and the NASA-bipolar methods. In Proceedings of the Human Factors Society 29th Annual Meeting (pp. 71-75). Santa Monica, CA: Human Factors Society.

Waugh, N.C. and Norman, D.A., 1965, Primary memory. Psychological Review, 72, 89-104.

Wickens, C.D., 1980, The structure of attentional resources. In R.S. Nickerson (Ed.), Attention and performance VIII (pp. 239-257). Hillsdale, NJ: Lawrence Erlbaum Associates.

Wickens, C.D., 1984, Engineering psychology and human performance. Columbus, OH: Charles E. Merrill Publishing Co.

Wickens, C.D., Born, D.G., and Allen, C.K., 1963, Proactive inhibition and item similarity in short-term memory. Journal of Verbal Learning and Verbal Behavior, 2, 440-445.

Wickens, C.D., Hyman, F., Dellinger, J., Taylor, H., and Meador, M., 1986, The Sternberg memory search task as an index of pilot workload. Ergonomics, 29 (11), 1371-1383.

Wickens, C.D., and Kessel, C., 1981, Failure detection in dynamic systems. In J. Rasmussen and W.B. Rouse (Eds.), Human detection and diagnosis of system failures (pp. 155-169), New York: Plenum Press.

Wickens, C.D., Moody, M.J., and Vidulich, M., 1985, Retrieval time as a function of memory set size, type of probes, and interference in recognition memory. Journal of Experimental Psychology: Learning, Memory, and Cognition, 11 (1), 154-164.

Wierwille, W.W., 1979, Physiological measures of aircrew workload. Human Factors, 21, 575-593.

Wierwille, W. W. and Casali, J. G., 1983, A validated rating scale for global mental workload measurement applications. In Proceedings of the Human Factors Society 27th Annual Meeting (pp. 129-133). Santa Monica, CA: Human Factors Society.

Wierwille, W. W. and Connor, S. A., 1983, Evaluation of 20 workload measures using a psychomotor task in a moving-base aircraft simulator. Human Factors, 25 (1), 1-16.

Wierwille, W. W. and Williges, R. C., 1978, Survey and analysis of operator workload assessment techniques (Tech Report S-78-101). Naval Air Test Center.

Williges, R. C. and Wierwille, W. W., 1979, Behavioral measures of aircrew mental workload. Human Factors, 21 (5), 549-574.

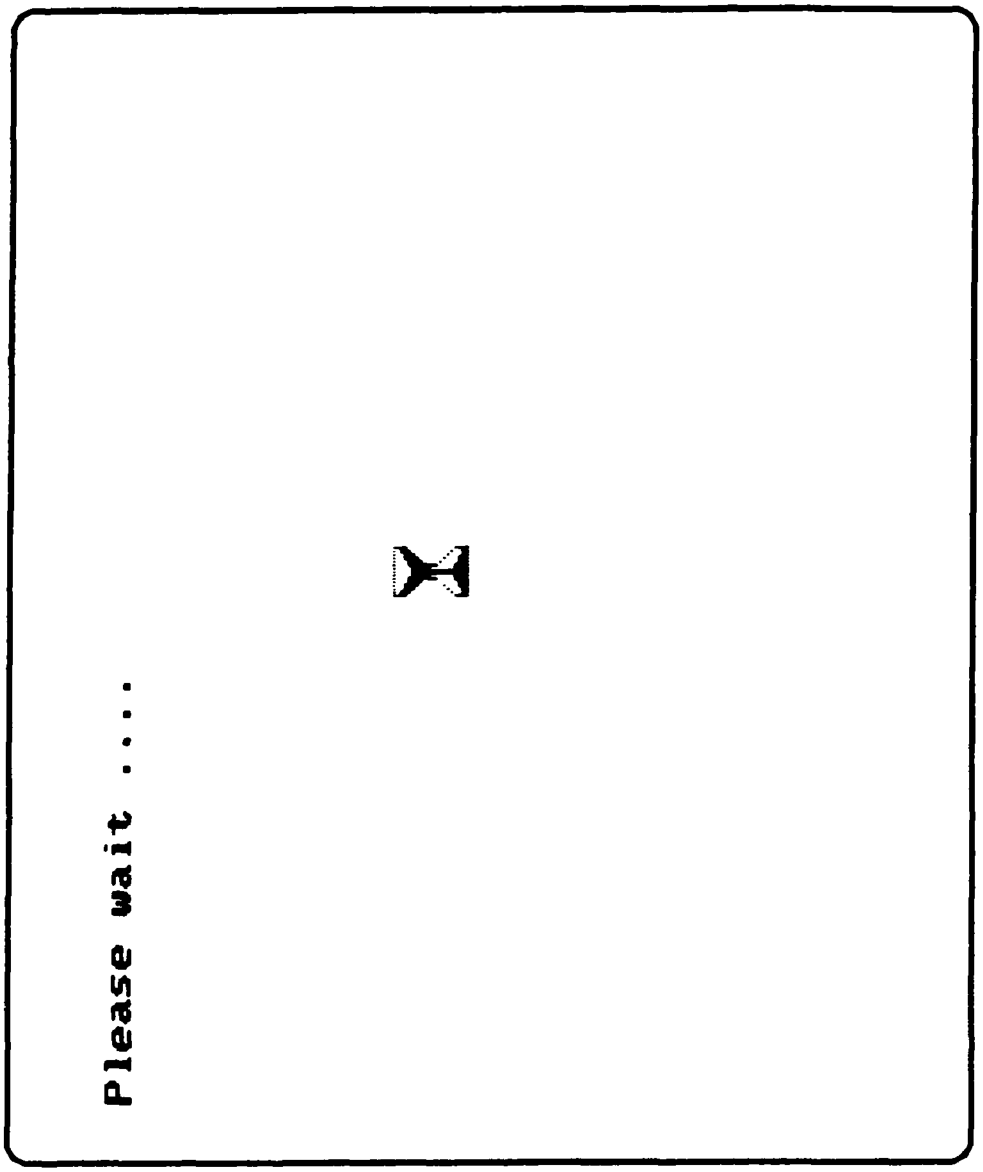
Wingfield, A. and Byrnes, D.L., 1981, The psychology of human memory. New York: Academic Press, Inc.

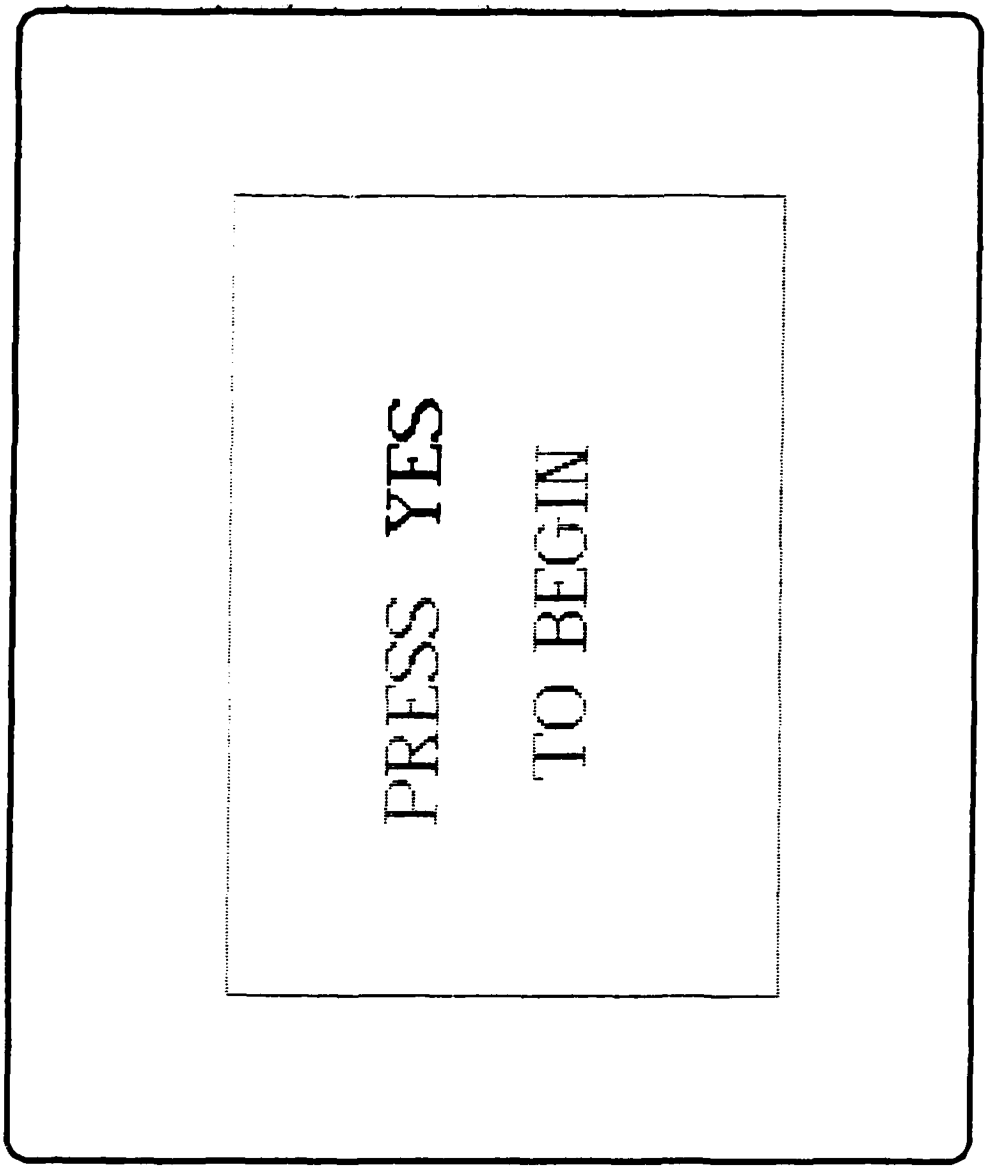
Yeh, Y. and Wickens, C. D., 1984, Why do performance and subjective measures dissociate? In Proceedings of the Human Factors Society 28th Annual Meeting (pp. 765-769). Santa Monica, CA: Human Factors Society.

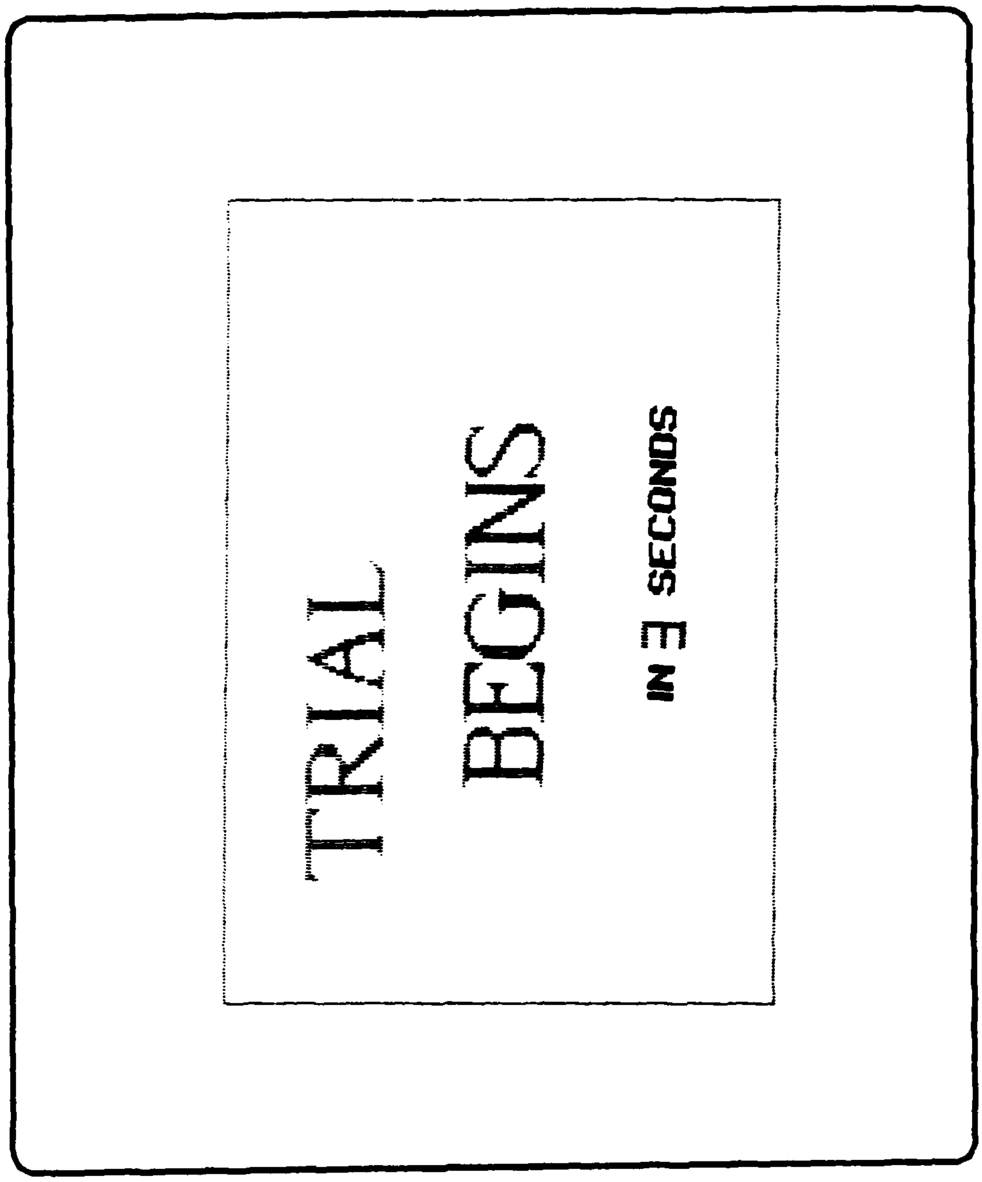
Yeh, Y. and Wickens, C. D., 1988, Dissociation of performance and subjective measures of workload. Human Factors, 30 (1), 111-120.

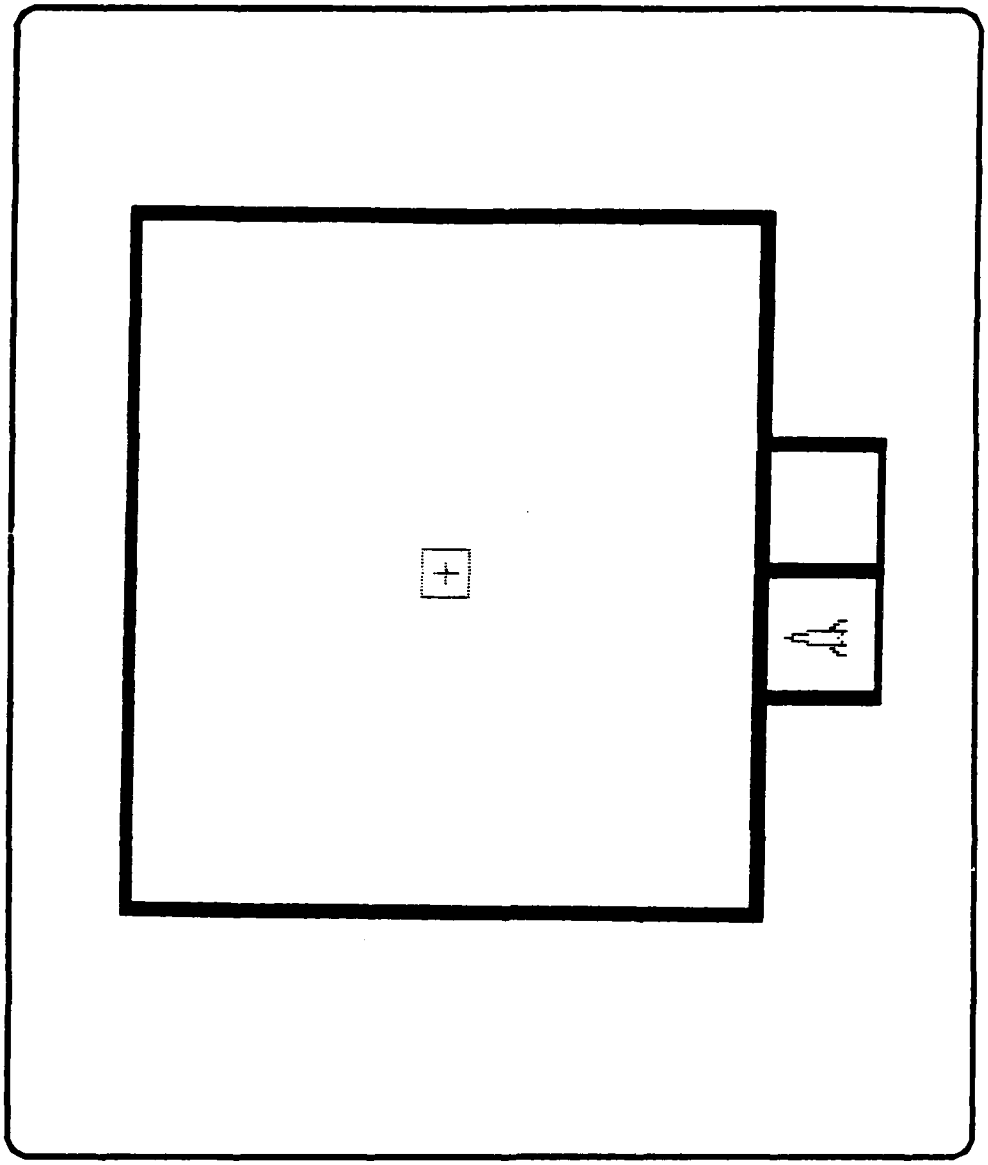
APPENDIX A

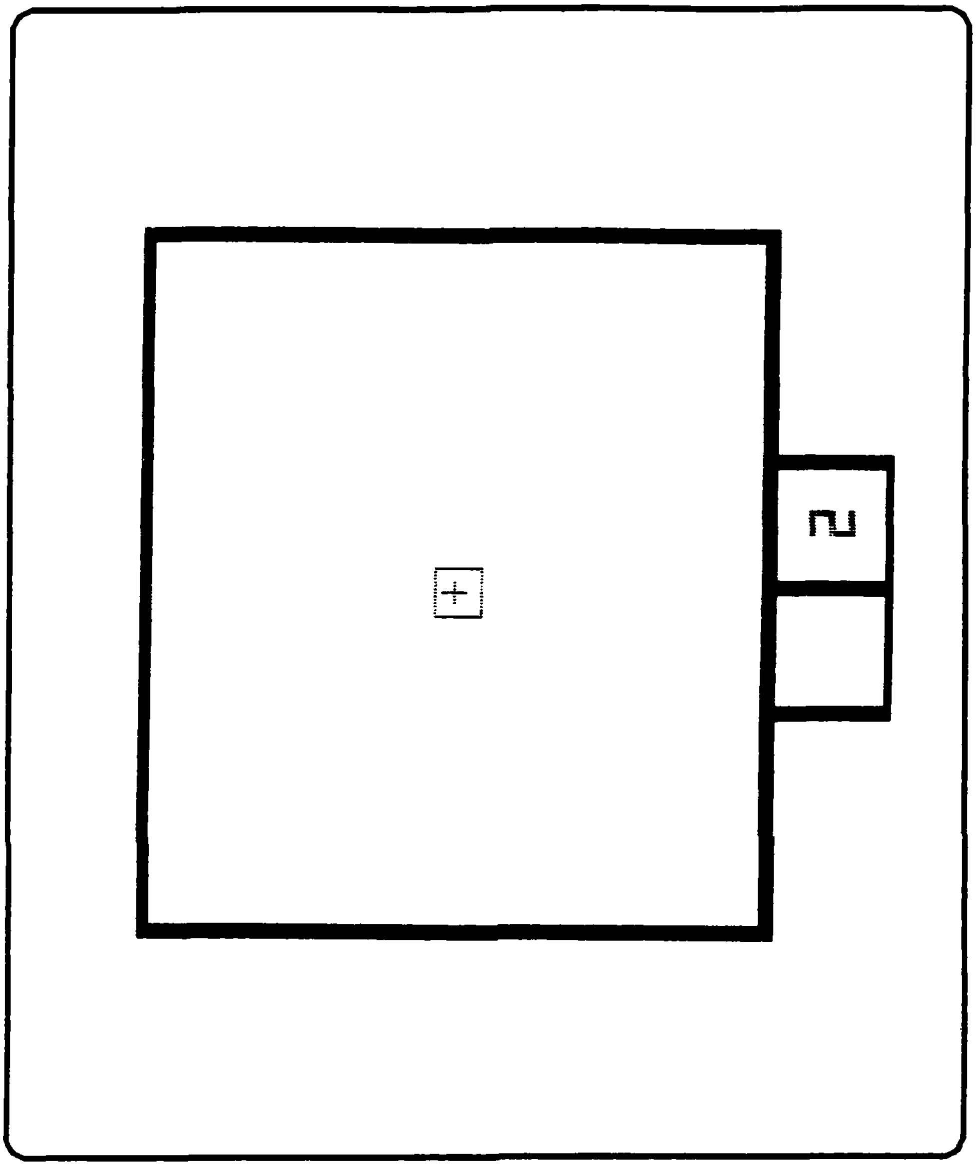
TASK SCREEN LAYOUTS

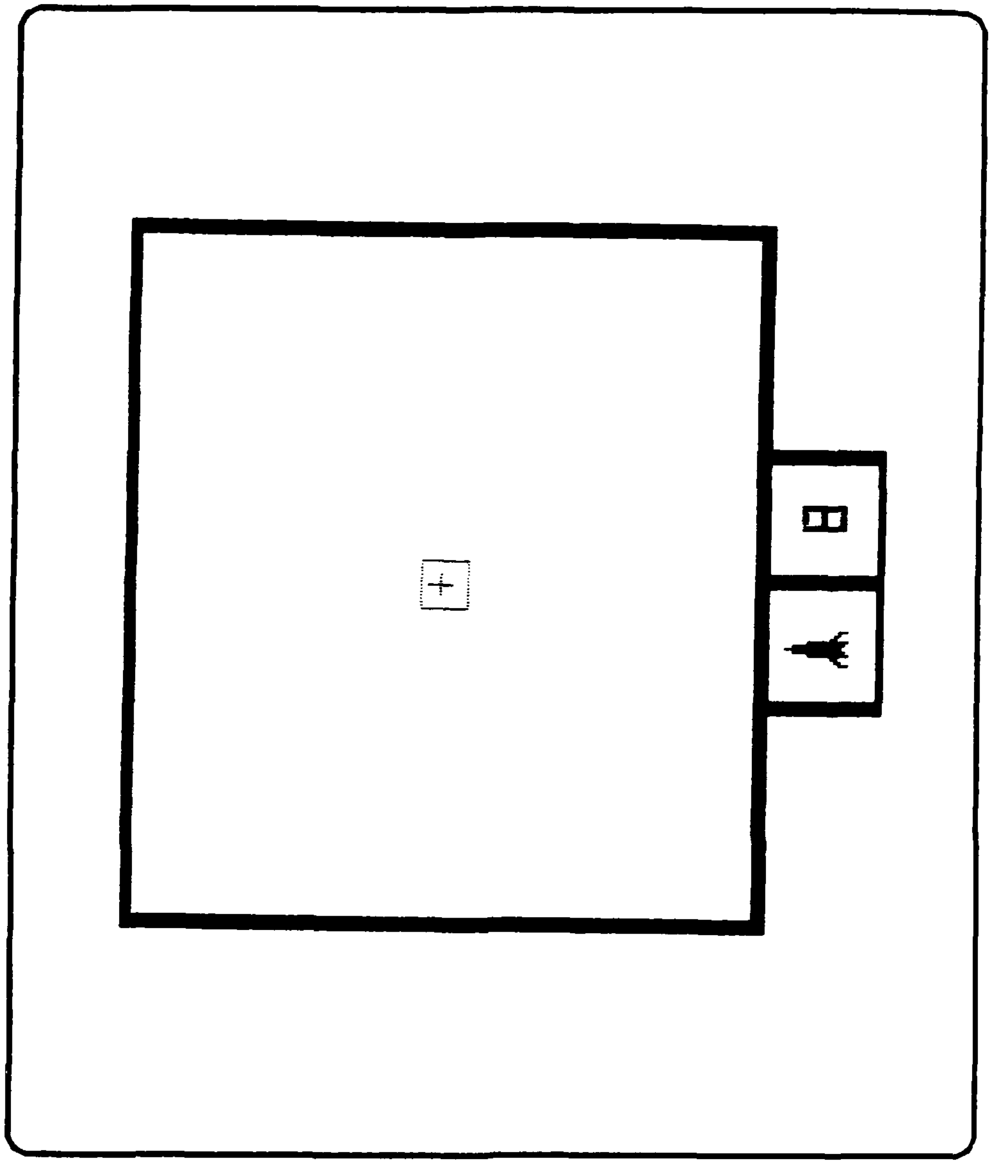



PRESS YES
TO BEGIN


TRIAL
BEGINS
IN 3 SECONDS







END OF TRIAL

ARE YOU

SATISFIED WITH

YOUR RATINGS ?

END OF PRACTICE

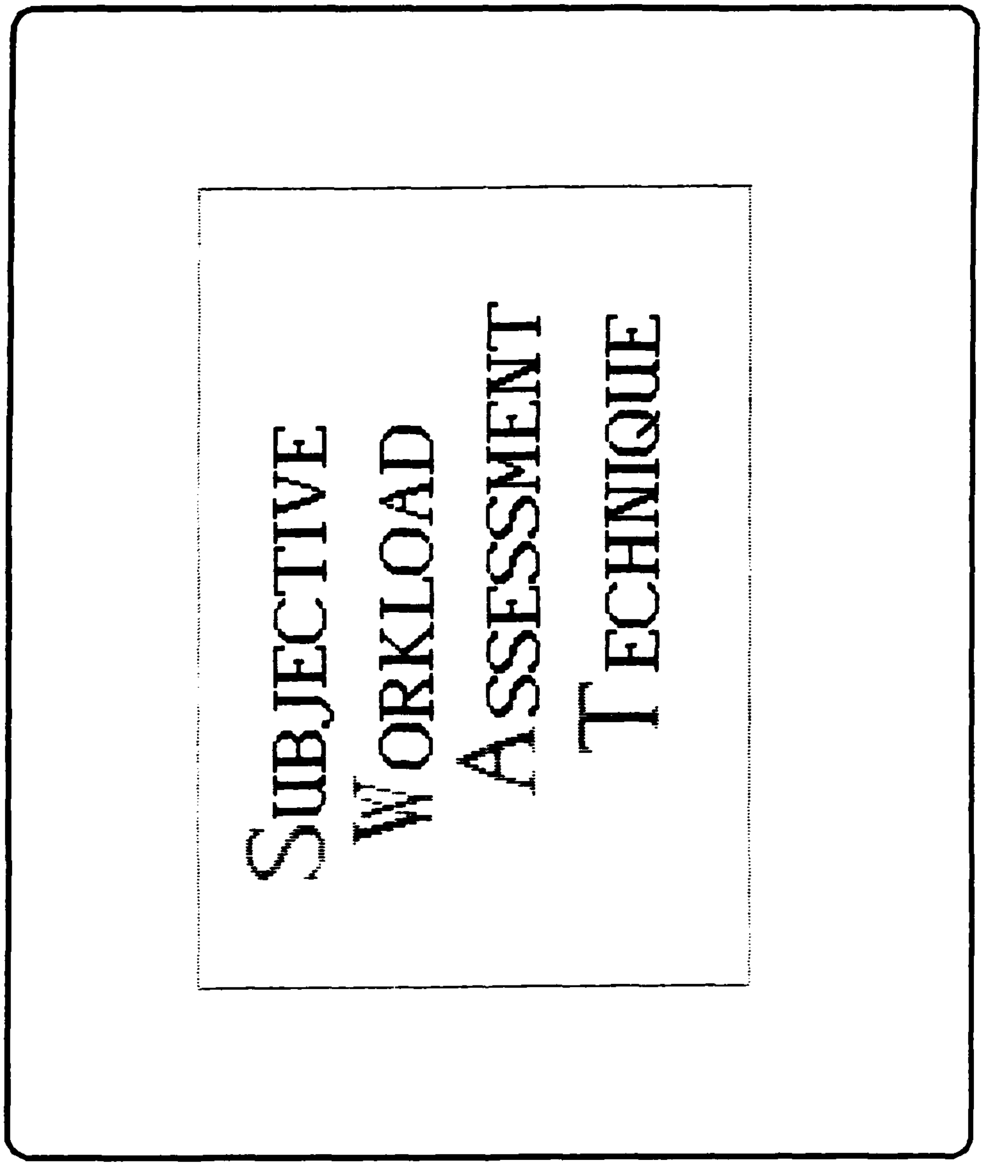
REST

PERIOD

END OF SESSION

APPENDIX B

SWAT SCREEN LAYOUTS



S
W
A
T

TIME LOAD	RATING
No or very few interruptions in the planning, execution, or monitoring of tasks. Spare time exists between many tasks.	1
Task planning, execution and monitoring are often interrupted. Little spare time. Tasks occasionally occur simultaneously.	2
Task planning, execution and monitoring are interrupted most of the time. No spare time. Tasks frequently occur simultaneously. Considerable difficulty.	3

S W A T

MENTAL EFFORT LOAD	RATING
Little conscious mental effort or planning required. Low task complexity such that tasks are often performed automatically.	1
Considerable conscious mental effort or planning required. Moderately high task complexity due to uncertainty, unpredictability, or unfamiliarity.	2
Extensive mental effort and skilled planning required. Very complex tasks demanding total attention.	3

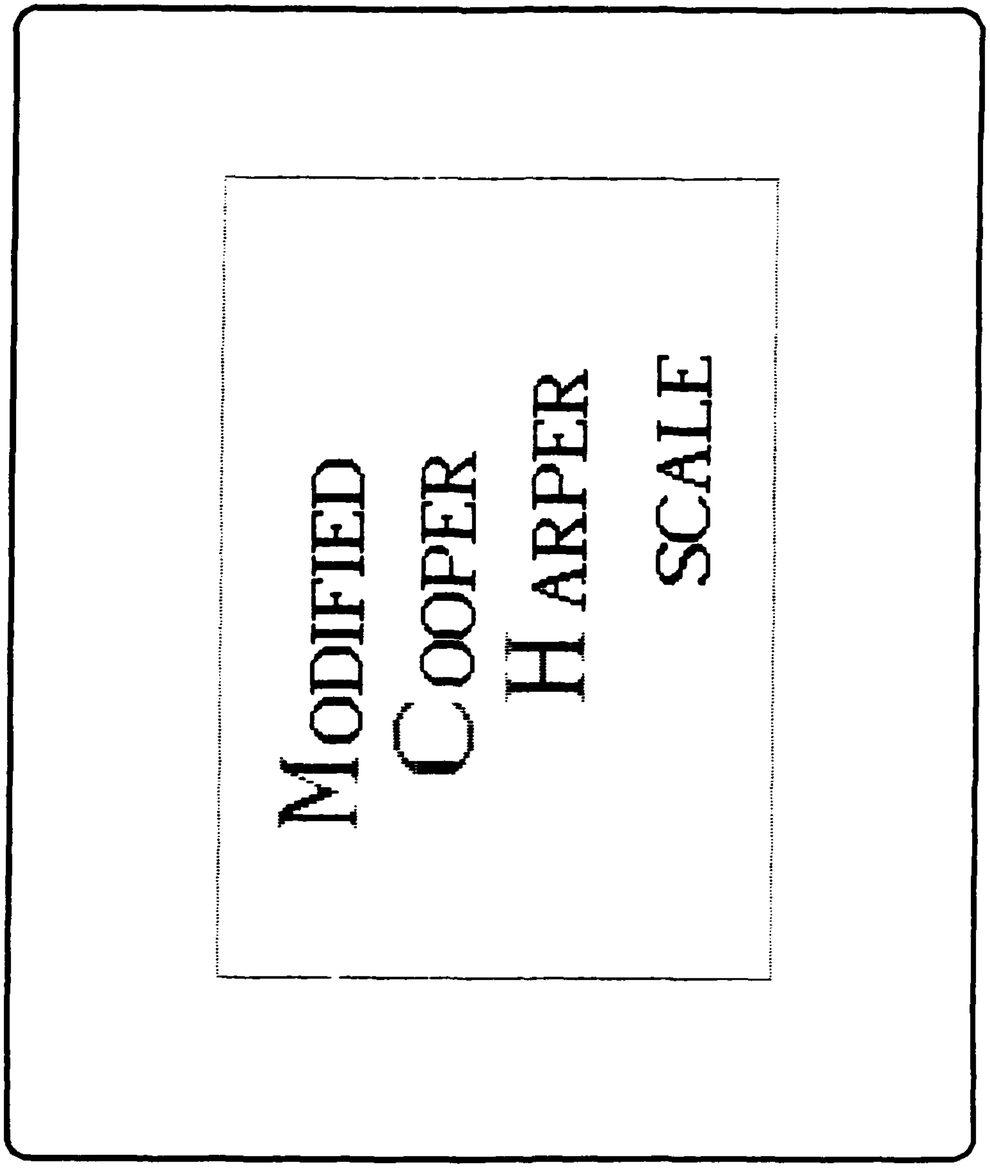
SWAT

PSYCHOLOGICAL STRESS LOAD

PSYCHOLOGICAL STRESS LOAD	RATING
Little risk, confusion, frustration, or anxiety exists and can be easily accommodated.	1
The degree of risk, confusion, frustration, or anxiety noticeably adds to workload and requires significant compensation to maintain adequate performance.	2
The level of risk, confusion, frustration, or anxiety greatly increases workload and requires tasks to be performed only with the highest level of determination.	3

APPENDIX C

MODIFIED COOPER-HARPER SCREEN LAYOUTS



Are errors
small and
inconsequential
?

DIFFICULTY LEVEL	OPERATOR DEMAND LEVEL	
Impossible	Instructed task cannot be accomplished reliably	⊕

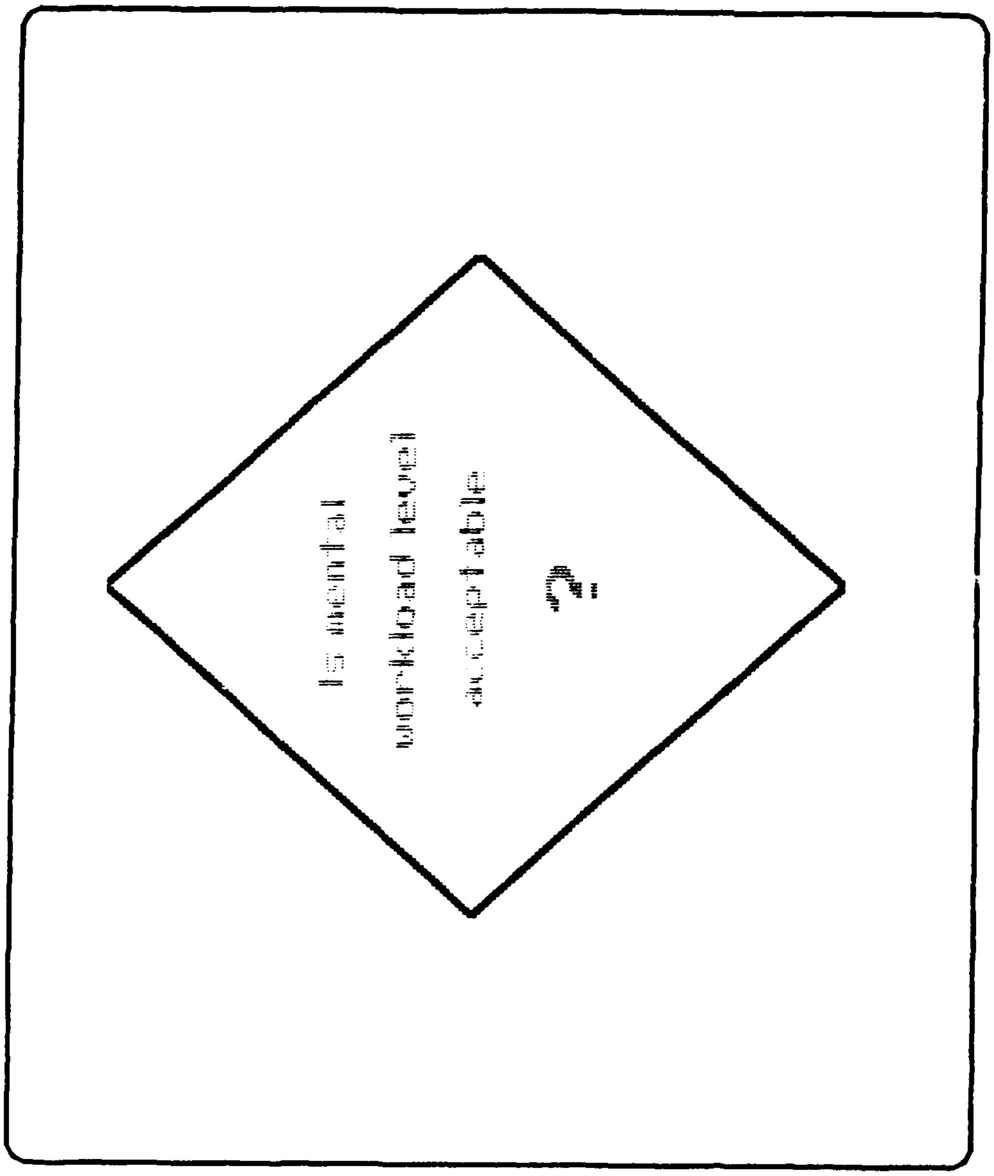
SYSTEM REDESIGN IS MANDATORY

Even
though
errors may be
large or frequent,
can instructed task be
accomplished most of
the time

?

DIFFICULTY LEVEL	OPERATOR DEMAND LEVEL	
Major Difficulty	Maximum operator mental effort is required to bring errors to moderate level	
Major Difficulty	Maximum operator mental effort is required to avoid large or numerous errors	
Major Difficulty	Intense operator mental effort is required to accomplish task, but frequent errors persist	⊕

SYSTEM REDESIGN IS STRONGLY RECOMMENDED


Is mental
workload level
acceptable
?

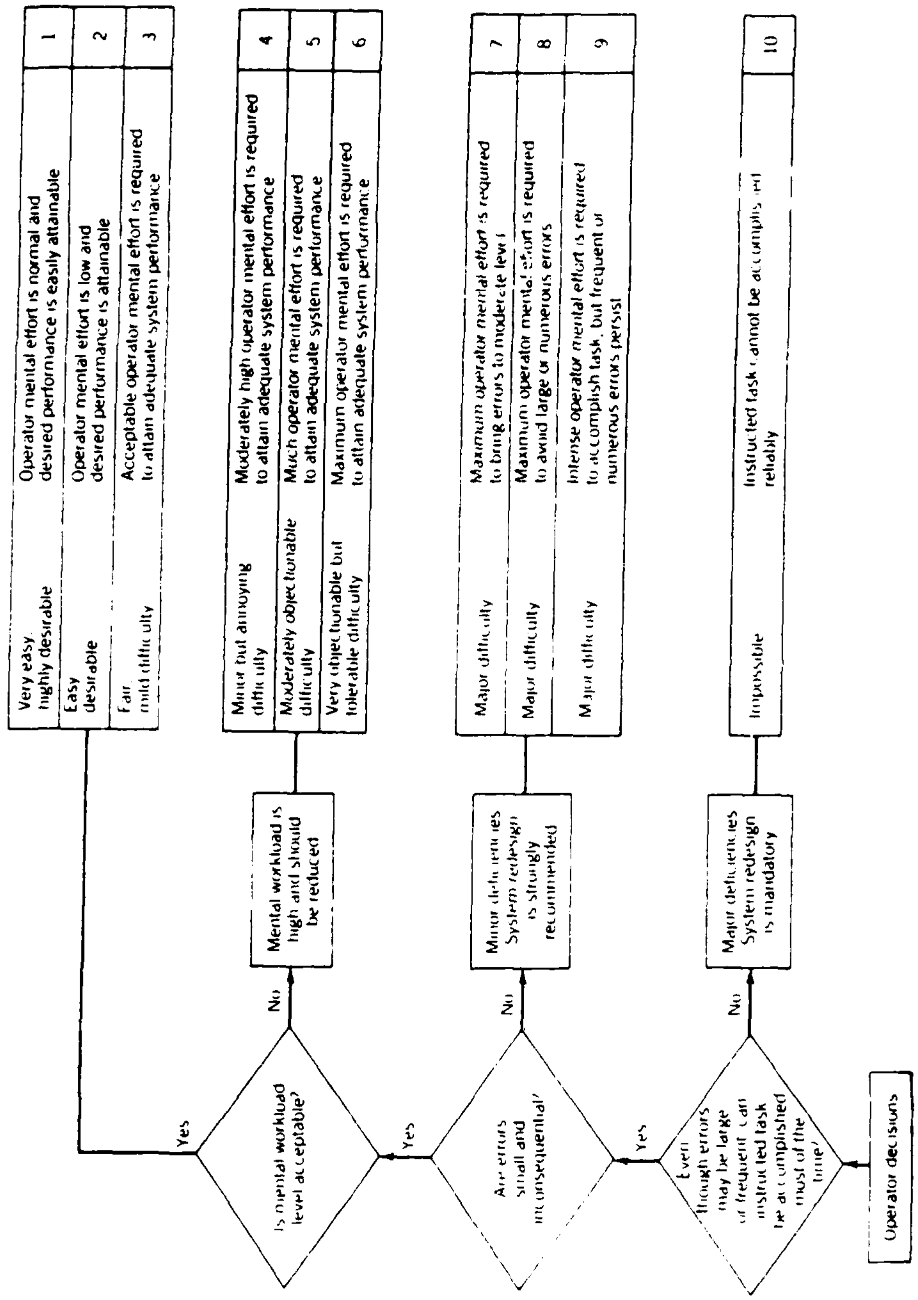
DIFFICULTY LEVEL	OPERATOR DEMAND LEVEL	
Minor but annoying difficulty	Moderately high operator mental effort is required	⊕
Moderately objectionable difficulty	High operator mental effort is required to attain adequate system performance	
Very objectionable but tolerable difficulty	Maximum operator mental effort is required to attain adequate system performance	

MENTAL WORKLOAD IS HIGH, SHOULD BE REDUCED

DIFFICULTY LEVEL	OPERATOR DEMAND LEVEL	
Very easy, highly desirable	Operator mental effort is minimal and desired performance is easily attainable	
Easy, desirable	Operator mental effort is low and desired performance is attainable	⊕
Fair, mild difficulty	Acceptable operator mental effort is required to attain adequate system performance	

APPENDIX D

MODIFIED COOPER-HARPER RATING SCALE



APPENDIX E

INSTRUCTIONS TO SUBJECTS

TRACKING INSTRUCTIONS

The sample display on the screen in front of you represents the simulated control environment that will be used in this experiment.

Target. The target that you will track is represented by the small green box in the middle of the screen. The target will remain stationary.

"Crosshair". Your tracking position is represented by the green crosshair ("+") near the target. Random forces will drive the crosshair away from the target. You can move the crosshair by pressing against the joystick in the desired direction.

Objective. Your task is to keep the crosshair as close to the center of the target as possible. You will complete one practice trial before the test trial. Each trial will last under two minutes.

Please operate the joystick with your preferred hand.

DO YOU HAVE ANY QUESTIONS ?

GENERAL EXPERIMENTAL INSTRUCTIONS

In this experiment, you will be asked to perform a number of tasks involving visual, auditory, and control interaction with a computer display. The display you will be using is an abstracted simulation of a control environment in which future U.S. aircrewmen may be working. Your tasks will include both primary and secondary responsibilities.

PRIMARY RESPONSIBILITIES

Target Tracking. As before, you are to keep the green crosshair ("+") as close to the center of the green target box as possible.

Missile Identification. Periodically, a missile symbol (like the one now on the screen) may appear in the lower left box on the screen. You are to report whether the missile is solid or hollow. A single, low frequency tone will warn you that the missile symbol will appear in 1/2 second. Respond accurately and quickly as follows:

> Press the "YES" key if the missile is **SOLID**.
> Press the "NO" key if the missile is **HOLLOW**.

Speed of response is important, but accuracy is more important. Please study the display to be sure you understand these symbols.

> *Remember*, you must continue to track the target. Make your missile identification *as quickly as possible without making an error* while you continue to track the target.

SECONDARY RESPONSIBILITIES

Digit List Monitoring. Since this is your secondary task, it should always take second place to your primary duties. After 10 and 30 seconds of tracking, a digit list consisting of 2, 3, 4, or 5 digits will be presented in the small red box to the right. You will be notified that a digit list is about to be presented by a single, high frequency tone (different than the tone used to signal a missile appearance).

The digit list will be presented one digit at a time, ranging from 0 to 9 with each digit remaining on the screen for one second. There will be no duplicate digits in a digit list. Following the presentation of the digit list and a brief pause, two rapid high frequency tones will be sounded. Following this signal, a single digit will appear in the box. Your task is to decide, as acurately and as quickly as possible, whether this digit was included in the digit list that you were just given.

> Press the "YES" key if the digit **IS** a member of the Digit Lisit.
> Press the "NO" key if the digit **IS NOT** a member of the Digit Lisit.

Remember, your primary responsibilities are target tracking and missile identification. Try your best to monitor these Digit Lists *without sacrificing your performance on your primary tasks.*

Tone Recognition. There will be three auditory signals used to provide you with specific information. These tones can occur at any time.

Single, low frequency tone ----------- missile symbol is about to appear.
Single, high frequency tone ---------- digit list is about to be presented.
Two. rapid, high frequency tones ---- the "test" digit will be presented.

AT THIS TIME, DO YOU HAVE ANY QUESTIONS CONCERNING YOUR TASK RESPONSIBILITIES OR THE VIDEO DISPLAY ?

PRACTICE SESSION

Please seat yourself in a comfortable position. Position your preferred hand gently on the joystick and your other hand on the switch box. The switch box requires only a light touch for data entry.

Practice a moment with these input devices.

This practice session is sequenced in a specific order of difficulty. Do not expect your experimental sessions to be predictable.

RATING SCALE INSTRUCTIONS

During the next three days, you will perform your tasks during a series of short (40 second) trials. At the end of each trial you will rate that trial using two scales (not always in the same order); the Modified Cooper-Harper scale and the Subjective Workload Assessment Technique (SWAT).

A separate computer screen for each of your decisions will be presented. Move the crosshair over your desired answer and push the "YES" key. Following each rating session, you will be asked if you are satisfied

with your ratings. If you have made a mistake or changed your mind, press the "NO" key and enter your ratings again.

RATING STRATEGIES

On all of your ratings, you will be evaluating the system for a general user population, not just yourself. You may assume you are an experienced member of that population and that your performance is typical of all other operators. Keep these points in mind.

First, give it your best effort. Be sure to try to perform the primary and secondary tasks *as instructed* and make all your evaluations within the context of the instructed tasks. Try to maintain adequate performance as specified for your tasks.

Second, rate the system. The rating scale is not a test of your personal skill. You should make the assumption that problems you encounter are not problems *you* created, but rather problems created by the system and the instructed tasks.

Third, rate consistently. Try to avoid biased ratings, e.g., being *overly critical* of a good system, or being *overly lenient* with a difficult system. Also, try not to overreact to small changes in the system. Thus, to avoid any problems, simply "tell it like it is" with your ratings.

MODIFIED COOPER-HARPER SCALE

DEFINITIONS

The terms used in the Modified Cooper-Harper Scale have specific meanings. It is important to begin with an understanding of these terms and how they apply in this experiment.

Primary Task refers to both the tracking task and the missile identification task.

Secondary Task refers to monitoring of Digit Lists.

System refers to the equipment used in performing the primary task. Together, you and the system make up the operator/system. For this experiment, the system includes the display screen, the joystick, the keypad, and the computer's sound generator.

Operator refers to *you*, the person performing the ratings. You will be operating the system and using the rating scale to quantify your experiences.

Errors can range from "small and inconsequential" to "large or frequent". In this experiment, errors

are any appreciable deviation from desired operator/system performance and can include any of the following: mistakes, incorrect actions or responses, blunders, omissions, and incompletions.

Mental Workload is the total mental effort required to operate the system, including the primary and secondary tasks. It includes such factors as level of attention, depth of thinking, and level of concentration required by the tasks.

RATING STEPS

The Modified Cooper-Harper scale (refer to handout) requires you to make a series of decisions. This scale follows a predetermined sequence designed to help you make consistent and accurate ratings. It is important that you follow this *complete* logic sequence for each of your ratings.

When you are rating a trial, go through the following steps, one computer screen at a time. Each screen will contain the appropriate portion of the Modified Cooper-Harper scale for your next decision.

Step 1. Were you able to accomplish both the primary and secondary tasks most of the time? If YES, go to Step 2. If NO, there is only one possible rating, and you are finished.

Step 2. Were the errors in performing either the primary or secondary tasks small and inconsequential? If YES, go to Step 3. If NO, rate the system by selecting the description that best summarizes the situation you experienced in that trial.

Step 3. Was the mental workload acceptable? If YES, go to Step 4. If NO, rate the system by selecting the most appropriate description.

Step 4. When the mental workload was acceptable as indicated in Step 3, again rate the system by selecting the description you deem most appropriate.

Remember, you may choose only one rating per trial, and the rating should be arrived at by following the scale's logic. You will always begin at the lower level and follow the logic path until you have decided on a rating. In particular, do not skip any steps in the logic. Otherwise, your rating may not be valid and reliable.

SUBJECTIVE WORKLOAD ASSESSMENT TECHNIQUE

DEFINITIONS

The Subjective Workload Assessment Technique (SWAT) uses three different dimensions to evalutate

mental workload. In general, these are as follows:

Time Load is the fraction of the total time that you are busy.
Mental Effort Load is the amount of attention and concentration required to perform the tasks.
Psychological Stress Load refers to conditions that produce confusion, frustration, anxiety, and/or risk during the performance of the task which causes a need for greater concentration and determination.

In this experiment, **Risk** refers to the risk of making any error.

INSTRUCTIONS

Like the modified Cooper-Harper scale, the SWAT requires you to rate the workload demands imposed by the primary and secondary tasks. This rating is accomplished in two phases.

Phase 1. This phase will be completed today. You will be given a stack of 27 index cards. Each card will contain a written description of each of the three workload dimensions: Time Load, Mental Effort Load, and Psychological Stress Load.

First, sort the cards into "high", "moderate", and "low" workload stacks with approximately nine cards in each stack. Make these decisions based on what **you** feel constitutes high, moderate, and low workload situations. There are no "right" or "wrong" sort orders and you **need not** have exactly nine cards in each pile.

Next, arrange each stack of cards into what you feel represents "lowest" to "highest" workload situations. Place the card you consider describing the "lowest workload situation" face up on the table. Maintaining the order you have selected, place each of the other cards (face up) on top of the first card. When you have placed the last card for that pile (the "highest workload situation") on the stack, secure the stack with a rubber band and place it on the table next to the appropriate card (marked "HIGH", "MEDIUM" and "LOW").

Phase 2. This will be part of your debriefing after each trial. You will be presented three computer screens for rating the SWAT workload dimensions. Each screen will contain three written descriptions. You are to select what you feel is an appropriate rating for each screen. As before, make your selection by moving the crosshair over the appropriate rating and pressing the "YES" key.

BEFORE YOU BEGIN THE FIRST PHASE OF YOUR SWAT RATING,
DO YOU HAVE ANY QUESTIONS ABOUT THESE INSTRUCTIONS?

SWAT INSTRUCTIONS FOR SECOND CARD SORT

DEFINITIONS

The Subjective Workload Assessment Technique (SWAT) uses three different dimensions to evalutate mental workload. In general, these are as follows:

Time Load is the fraction of the total time that you are busy.
Mental Effort Load is the amount of attention and concentration required to perform the tasks.
Psychological Stress Load refers to conditions that produce confusion, frustration, anxiety, and/or risk during the performance of the task which causes a need for greater concentration and determination.

In this experiment, **Risk** refers to the risk of making any error.

INSTRUCTIONS

You will be given a stack of 27 index cards. Each card will contain a written description of each of the three workload dimensions: Time Load, Mental Effort Load, and Psychological Stress Load. Each of these three dimensions contibutes in some way to workload. Together, the combination of the three descriptions on each card describes an imaginary "workload situation."

First, sort the cards into "high", "moderate", and "low" workload stacks with approximately nine cards in each stack. Make these decisions based on what **you** feel constitutes high, moderate, and low workload situations. There are no "right" or "wrong" sort orders and you **need not** have exactly nine cards in each pile.

Next, arrange each stack of cards into what you feel represents "lowest" to "highest" workload situations. Place the card you consider describing the "lowest workload situation" face up on the table. Maintaining the order you have selected, place each of the other cards (face up) on top of the first card.

When you have placed the last card for each pile (the "highest workload situation") on top of each stack, you may combine the three stacks into one pile (with the LOW stack on bottom and the HIGH stack on top). You should then look through the entire stack again to make sure the cards are ordered in what you feel to be the "lowest" to "highest" workload order.

It is important that you give your best effort in following these instructions. You should have plenty of time in which to complete the card sort (up to one hour). In addition, you are allowed to ask any questions you may have about these instructions, now or during the card sort.

DAILY REVIEW INSTRUCTIONS

Before beginning today's experiment, you will have a short practice session to refresh your memory of the system. First, please review the instructions.

1. **Target Tracking.** Your first primary task is to keep the crosshair as close to the center of the target as possible at all times.

2. **Missile Identification.** Your second primary task is to report whether or not the random missile symbol is solid or hollow. Response speed is important, however accuracy is more important.

Press the "YES" key if the missile is SOLID.

Press the "NO" key if the missile is HOLLOW.

Remember, make your choice *as quickly as possible without making an error.*
Also, continue tracking the target at all times.

3. **Digit List Monitoring.** Twice during each trial, a digit list (with 2, 3, 4, or 5 digits) will be presented. After seeing this list you are to identify whether or not the following digit is part of that list.

Press the "YES" key if the digit IS part of the digit list.

Press the "NO" key if the digit IS NOT part of the digit list.

Again, it is important to respond as quickly as possible, but more important that you not make a mistake.

4. **Tone Recognition.** Three auditory signals are used to provide specific information. These tones can occur at any time.

Single low frequency tone ------- a missile symbol is about to appear.

Single high frequency tone ------- a digit list is about to be presented.

Two brief, high frequency tones --- a "test"digit will be presented.

Remember, your primary responsibility is with your tracking and symbol identification tasks. Try your best to monitor these digit lists, *but without sacrificing your performance on primary tasks.*

5. **Debriefing Each Trial.** After each trial (approximately every minute), you will rate that trial using both the Modified Cooper-Harper scale for workload and the Subjective Workload Assessment Technique (SWAT).

a. **Modified Cooper-Harper Scale.** Arrive at your rating by following the logic path until you have decided on a rating for the trial you just finished. Do not skip any steps. Otherwise, your rating may not be valid and reliable.

b. **Subjective Workload Assessment Technique (SWAT).** SWAT uses three mental workload dimensions: Time Load, Mental Effort Load, and Psychological Stress Load. You will be asked to rate each of these dimensions by selecting the written description most closely representing your experiences during that trial.

To avoid any problems with either of these scales, simply **"tell it like it is"** in making your ratings.

BEFORE YOU BEGIN YOUR PRACTICE SESSION, DO YOU HAVE ANY QUESTIONS ABOUT THESE INSTRUCTIONS OR YOUR TASKS ?

APPENDIX F

PARTICIPANT'S INFORMED CONSENT FORMS

INFORMED CONSENT FORM FOR SCREENING PROCEDURE

Before you are asked to participate as a subject in the research project, we ask that you complete a brief screening procedure. The purpose of this procedure is to determine whether your vision and manual coordination meet the criteria we have established for participating in this experiment. These are not professional tests; therefore, the results should not be considered accurate descriptions of your performance capabilities. In particular, a professional eye doctor should be consulted for an accurate description of your vision.

This screening procedure consists of three parts. If your vision meets the criteria for part 1, you will proceed to part 2, also a vision test. If you pass both vision examinations, you will perform a brief manual control task, similar to that which you would perform in the experiment. This screening procedure is expected to take no longer than 20 minutes. If you pass the screening procedure, you will be asked to participate in the experiment. You will be paid for participating in the screening procedure and if you participate in the experiment, you will be paid for your time spent in the experiment.

As a subject in this screening procedure you are entitled to certain rights:

1.) You may withdraw from participation in this procedure at any time you wish without penalty. However, if you do so, you will not be asked to participate in the experiment.

2.) The principal investigator of this project or his associates will answer any questions you may have concerning this procedure, and you should not sign this consent form until you are satisfied that you understand all the terms involved.

3.) The IEOR research team members on this project include:
William F. Reinhart, Graduate Student;
Craig Dye, Graduate Student;
Carita Glynn, Graduate Student;
Mark Takahama, Graduate Student; and
Dr. Harry L. Snyder, Faculty Member.

4.) The data collected during your participation will be treated with confidentiality and used soley for

purposes of screening for the research project.

If you have further questions about your rights as a participant, you may contact Mr. Charles D. Waring, Chairman of the Institutional Review Board at Virginia Tech.

Your signature below indicates that you have read this document in its entirety, that your questions have been answered, and that you consent to participate in the screening procedure described.

The faculty and graduate students involved in this research appreciate your participation.

________________________ ________________________

Signature Telephone number

____

Printed name

Displays and Controls Laboratory
Industrial Engineering and Operations Research
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061
(703) 961-5499

INFORMED CONSENT FORM FOR THE EXPERIMENT

You are being asked to participate as a subject in a research project. The purpose of this experiment is to examine performance and mental workload under varying levels of task complexity. The tasks you will be asked to perform are presented in a video game context and will involve manual joystick control, visual target identification and response, and remembering short lists of digits. You will also be asked to rate the importance of various task factors in a card sorting procedure.

The experiment is expected to last 4 consecutive days, for a maximum total of 5 hours. If you decide to participate, you will be paid $5.00 per hour for the time you spend in the laboratory, or $25.00 for completion of the experiment, whichever is greater. Payment will be made upon completion of your participation.

As a subject in this experiment you are entitled to certain rights:

1.) You may withdraw from participation in this research project at any time you wish without penalty. However, if you do so, you will only be paid for the time which you actually spend participating in the experiment.

2.) The principal investigator of this project or his associates will answer any questions you may have concerning this research, and you should not sign this consent form until you are satisfied that you understand all the terms involved. However, in cases where experimental details may affect the outcome of the experiment, the researcher may delay a complete disclosure until you have completed the experiment.

3.) The IEOR research team members on this project include:

William F. Reinhart, Graduate Student;
Craig Dye, Graduate Student;
Carita Glynn, Graduate Student;
Mark Takahama, Graduate Student; and
Dr. Harry L. Snyder, Faculty Member.

4.) If you wish to receive a summary of the results of this research, please include your address (where you expect to be living three months from now) with your signature below. Please do so only if you are truly interested in seeing the results.

5.) The data collected during your participation will be treated with anonymity. After you have participated, your name will be separated from your data. For this reason, if you wish to withdraw your data from our analyses, you must notify the experimenter immediately after your participation is complete.

If you have further questions about your rights as a participant, you may contact Mr. Charles D. Waring, Chairman of the Institutional Review Board at Virginia Tech.

Your signature below indicates that you have read this document in its entirety, that your questions have been answered, and that you consent to participate in the study described.

The faculty and graduate students involved in this research appreciate your participation.

____________________ ____________________

Signature Address

____________________ ____________________

Printed name

____

Displays and Controls Laboratory
Industrial Engineering and Operations Research
Virginia Polytechnic Institute and State University
Blacksburg, Virginia 24061
(703) 961-5499